Universität Ulm
Abteilung Zahlentheorie/Wahrscheinlichkeitstheorie

Dependencies of Extreme Events in Finance

Modelling, Statistics, and Data Analysis

Dissertation
zur Erlangung des Doktorgrades
Dr. rer. nat.
der Fakultät für Mathematik und Wirtschaftswissenschaften
der Universität Ulm

vorgelegt von

Dipl.-Math. oec. Rafael Schmidt

Ulm, im Mai 2003

Amtierender Dekan: Prof. Dr. Dieter Beschorner

1. Gutachter: Prof. Dr. Ulrich Stadtmüller, Universität Ulm
2. Gutachter: Prof. Dr. Rüdiger Kiesel, Universität Ulm
3. Gutachter: Prof. Dr. Nick Bingham, Brunel University

Tag der Promotion: 24. Juli 2003

Preface

In the beginning of the present thesis I would like to express my deepest gratitude to Professor Ulrich Stadtmüller for his excellent guidance and supervision during the entire phase of my doctorate. He filled me with enthusiasm in many fields of mathematics and is an excellent teacher.

I am also very grateful to Professors Nick Bingham and Rüdiger Kiesel for many helpful discussions and comments. Especially I would like to thank them for giving me the opportunity to spend an extremely fruitful research time at the University of London.

Moreover, I am very much indebted to Professor Gholamreza Nakhaeizadeh, the Department of Information Mining and my colleagues in the Risk Group at the Research Center Ulm of DaimlerChrysler AG, and the project-team "Accord" at DaimlerChrysler Services in Berlin. Many theoretical results of the present thesis are induced by practical questions and problems from the field of risk management which have been revealed during a collaboration between the Department of Number Theory and Probability Theory at the University of Ulm and the Department of Information Mining at the DaimlerChrysler AG, Research Center, in Ulm under the supervision of Professors Gholamreza Nakhaeizadeh and Ulrich Stadtmüller. Further, the theoretical results derived in the course of this thesis have strongly influenced several practical solutions which have been established during the above collaboration. The practical experiences in the fields of data mining and risk management, I have gained during the years at the Research Center, have broadened my horizons and are indispensable for my future career.

My special gratitude belongs to the rector of the University of Ulm, Professor Hans Wolff, for his support, valuable advice and many fruitful discussions. I would also like to show my appreciation to the Head of the Department of Number Theory and Probability Theory at the University of Ulm, Professor Helmut Maier, for offering me a stimulating research environment.

Further, I thank Professor Ted Cox for introducing me to the world of stochastic processes during a one-year stay as exchange student at Syracuse University.

I also extend my thanks to my colleagues Dr. Tomas Hrycej, Dr. Hartmut Lanzinger, Gabriel Frahm, and Markus Junker for helping me in my mathematical understanding and suggesting interesting problems.

Finally, I am deeply indebted to my parents Maria and Volker, to my sister Evelyn, and to Sandra for supporting, motivating and encouraging me during all stages of my life.

The financial support of DaimlerChrysler AG Research and Technology, the Departments of Number Theory/Probability Theory and Financial Mathematics at the University of Ulm, and the DAAD for the research leading to this thesis is gratefully acknowledged.

Ulm, 10th May 2003

Contents

Chapter 1

Introduction and summary

Dependencies between financial asset-returns have significantly increased during recent time periods in almost all international markets. This phenomenon is a direct consequence of globalization and relaxed market regulation in finance and insurance industry. Especially during bear markets many empirical surveys like Karolyi and Stulz (1996), Longin and Solnik (2001), and Campbell, Koedijk, and Kofman (2002) show evidence of increasing dependencies between financial asset-returns. However, increasing dependencies strongly impact the companies profit contributions and may weaken the financial stability of entire industrial sectors. Typically, risk managers pursue diversification strategies by analyzing and utilizing positive and negative correlations between various asset-returns in order to cut one's losses due to *market* or *credit risk* and to increase the (risk-adjusted) returns. Though, diversification strategies become less effective or may break down if the financial markets fall simultaneously during bear markets or market crashes. In our context, market risk relates to the risk of losses in on- and off-balance sheet positions arising from movements in market prices of financial assets whereas credit risk refers to possible losses in the credit portfolio. Most banks distinguish between two conceptual definitions of credit loss (Basel Committee on Banking Supervision (1999)): the default-mode paradigm, in which a credit loss arises only if a borrower defaults within the planning horizon, and the mark-to-market (or more accurately, mark to model) paradigm, in which credit deterioration short of default is also incorporated.

The determination of an optimal diversification strategy requires a mathematical portfolio model whose main output consists of a distribution function describing the portfolio's return (or loss). Standard portfolio selection is usually based on the Markowitz mean-variance theory of risk and return, from 1952 on, and the Sharpe-Lintner-Mossin capital asset pricing model (CAPM) of 1964-66. For example the famous Markowitz portfolio model (cf. Markowitz (1987)) provides an investment strategy by utilizing the following mean(μ)-covariance(Σ) optimization approach. Consider an n-dimensional random vector describing the probabilistic behavior of asset-returns within a certain portfolio. Let α denote the vector of portfolio-weights, B denote the maximal budget, and μ and Σ denote the mean vector and covariance matrix of the latter asset-return vector. The Markowitz optimization problem (without short-selling) is now formulated by

$$\sup_{\alpha} \alpha'\mu \qquad (s.t.) \begin{cases} \alpha'\Sigma\alpha \leq B, \\ \sum_{i=1}^{n} \alpha_i = 1, \\ \alpha_i \geq 0, \ i = 1, \ldots, n. \end{cases} \tag{1.1}$$

In particular, the *dependence structure* of the asset returns is solely described by the covariance matrix Σ. This suffices in case the asset-return distributions are multivariate normal. However, if we leave the Gaussian world, a sole consideration of the covariance matrix often explains the dependence structure in a quite unsatisfying way. Pitfalls like a non-existing covariance or zero covariance for dependent random variables may occur. Especially the dependence structure of *extreme events* is usually poorly or incorrectly described by the covariance matrix. Extreme events refer, for example, to extraordinary claims to insurance companies, crashes of equity markets, or extreme losses in credit portfolios due to borrower defaults. Hence, extreme events occur rarely, ergo, only few extreme observations are available; but probabilities and dependence structures have to be assigned to extreme events due to their economic impact. Particularly in credit risk management, where dependencies between large credit losses are of principal interest, measures of *dependencies of extreme events* are needed in order to support appropriate asset-allocation strategies. Embrechts, McNeil, and Straumann (1999) provide further reasoning why traditional dependence measures such as covariance and correlation are not always suited for a proper understanding of dependencies in financial markets.

The theory of *copulae* (cf. Joe (1997), Nelsen (1999)) provides an important class of functions, namely copula functions or copulae, which describe the dependence structure of general multidimensional random vectors in a quite complete manner. In particular, copulae allow to study the dependence structure of random vectors irrespective of their marginal distributions and copulae can be used to build multivariate distribution functions in two steps: First, model the dependence structure with some copula and second, plug in appropriate margins (cf. Schweizer (1991)). Recently, the theory of copulae has been introduced to mathematical finance (Bouyé, Durrleman, Nikeghbali, Riboulet, and Roncalli (2000), Embrechts, Lindskog, and McNeil (2001), Junker and May (2002)) e.g. in order to study the dependence structure of asset returns irrespective of the marginal asset-return distributions; and as a starting point to construct multivariate distributions for asset-return vectors or to develop powerful simulation algorithms. Difficulties arising in the context of financial applications are mostly related to the right choice of copulae in higher dimensions. Especially, robust estimation and fast simulation methods are usually rare. In this thesis we investigate and explore the problem of appropriate dependence structures and dependence measures, and put the main emphasis on the dependence modelling of extreme events.

It has been already mentioned that dependencies of extreme events like extreme asset-returns (in short: *extremal dependencies*) play an important role in market and credit risk management. According to Ong (1999), the primary issue risk managers have always been interested in, is assessing the *size* - more than the *frequency* - of losses. The presumable most well-known risk measure called the Value-at-Risk (VaR) (describes the amount of extreme portfolio loss which is exceeded only with a certain small probability) depends strongly on the dependence structure of extreme events (for background on VaR measures we refer to Dowd (1998) and Jorion (2000)).

Summary and outline of the thesis

The main aim of this dissertation concerns the modelling and statistical investigation of dependencies between extreme events. Several simulation studies and data analysis emphasis the applicability of the obtained concepts and results to finance. Further, various integrative distribution models are proposed to describe more realistic models for financial assets.

Multivariate extreme-value theory (EVT) (in particular the multivariate extreme value distribution) is the natural choice for modelling multidimensional extreme events. However, the dependence structure of multivariate extreme value distributions and their relationship to the original distribution functions are mostly unknown to mathematicians and practitioners which are not that familiar with multivariate EVT. One aim of this thesis is to introduce so-called *tail copulae* which analyze the dependence structure of extreme events and incorporate many well-known properties of copulae. However, the class of tail copulae is quite large which might be a possible drawback for tail copulae to serve as general measures of extremal dependence in practical applications. Note that the correlation has become so popular to practitioners exactly because the dependence structure of a bivariate distribution is described by merely one value. Therefore, a major aim of this thesis is to put forward the concept of *tail dependence* which measures extremal dependence of a bivariate distribution function via two values: the so-called *lower and upper tail-dependence coefficient* (TDC) (cf. Joe (1997)). Important characterization and estimation methods will be developed for a large number of distributions and copulae. A strong emphasis is put on the class of so-called *multivariate elliptically contoured* distributions and copulae. This class contains the normal distributions and copulae, and inherits many useful analytical and statistical properties from the latter. For example we show that simulation and estimation algorithms become a relief in higher dimensions. Especially within financial applications, elliptically contoured distributions turn out to be quite useful and often appropriate as a substitute of the normal distribution.

Complementary to the exploration of extremal dependence structures, several types of integrative distribution models are investigated whose overall dependence structures (copulae) are suited for modelling of financial data, in particular, for asset-return modelling. At this point we clarify how to model asset returns. Statistically, only distinct asset prices $X_m,\ m = 1, \ldots, T$, are observable which correspond to equidistant time points. The *relative asset return* $\tilde{R}_m$ at time point m is defined by

$$\tilde{R}_m := \frac{X_m - X_{m-1}}{X_{m-1}}. \tag{1.2}$$

Nevertheless most authors prefer the log-*return*, instead of the relative return, which is given by

$$R_m := \log(X_m) - \log(X_{m-1}) = \log(1 + \tilde{R}_m). \tag{1.3}$$

A nice property of this definition is that the log-return over k periods can be written as the sum $R_m + \cdots + R_{m+k-1} = \log(X_{m+k-1}) - \log(X_{m-1})$. The difference between R_m and $\tilde{R}_m$ is numerically negligible since $\tilde{R}_m - R_m = \frac{1}{2}R_m^2 + \frac{1}{6}R_m^3 + \ldots$, and R_m is typically of order less than 10^{-2}. However, the latter approximation might not be satisfactory in the context of extreme events.

Modern market and credit-risk models primarily utilize integrative distribution models like multivariate normal distributions (see RiskMetrics (1995), CreditMetrics (1998)), multivariate generalized hyperbolic distributions (see Bingham and Kiesel (2001a), Eberlein (2001)), or multivariate alpha-stable distributions (see Rachev and Mittnik (2000)) for asset-return modelling. Various empirical surveys reveal (see also Section 2.1.3 and Chapter 5 of the present thesis) that, especially within the field of risk management, normal distributions are not always suited due to their symmetry and thin tails. On the other hand, multivariate generalized hyperbolic distributions and multivariate alpha-stable distributions usually lead to difficulties regarding the estimation of parameters, the generation of multivariate random numbers, and the modelling of dependence structures. To overcome these difficulties, this thesis proposes several models of integrative distributions which are either based on a multivariate elliptically contoured distribution (of a semi-parametric type) or they are based on a multivariate affine generalized-hyperbolic distribution. From a practical point of view, both distributions possess appealing dependence structures and useful statistical properties in higher dimensions.

Chapter 2 provides the reader with the necessary background on extreme-value theory and introduces the following theoretical concepts: copulae, tail copulae, tail dependence, regular variation, O-regular variation, and multivariate regular variation. In addition, we give a short guideline containing several exploratory visual aids to detect extremal dependencies in financial time series.

In Chapter 3, the concept of tail dependence is investigated for various distributions and copulae. Sections 3.1 and 3.2 establish general characterizations for spherically and elliptically contoured distributions which possess the so-called tail-dependence property. Theorems 3.1.4 and 3.2.2 provide the main results. Here we prove that spherically and elliptically contoured distributions are tail dependent if the tail of their generating random variable or density generator is regularly varying, i.e., the tail can be essentially approximated by a power function. Further, we give a necessary condition for tail dependence which is somewhat weaker than regular variation of the latter tail, namely, an O-regularly varying tail is required. O-regularly varying functions have growth (or decay) bounded between powers. In addition we derive an explicit formula (cf. formula (3.18)) for the so-called tail-dependence coefficient of elliptically contoured distributions. Up to our knowledge, the latter characterizations were published in Schmidt (2002) and seem to be new. The sufficient condition of tail dependence for elliptically contoured distributions has recently been obtained in Hult and Lindskog (2002) which, however, utilize different proof techniques. Besides, we provide several interesting results concerning the relationship between regularly (O-regularly) varying generating variates, density generators, and marginal distribution functions of spherically and elliptically contoured distributions. In addition, we check whether well-known examples of elliptically contoured distributions such as the multivariate normal, t, logistic, and symmetric generalized hyperbolic distributions are tail dependent or tail independent.

In Section 3.3 we embed the concept of tail dependence for elliptical copulae and distributions in multivariate EVT. In particular, Theorems 3.3.2 and 3.3.5 give sufficient conditions for elliptical copulae and distributions to be in the domain of attraction of an extreme value distribution. Finally, Section 3.4 provides several characterizations of tail dependence for Archimedean copulae and Section 3.5 explores the concept of tail dependence for several other copulae.

Chapter 4 is devoted to the estimation of extremal dependencies. Section 4.1 deals with the concept of tail copulae which describe the dependence structure in the tail of multivariate distributions and forms a generalization of the concept of tail dependence. Several nonparametric estimators of the tail copula are introduced in Section 4.2. Further, we propose a nonparametric estimator for the lower and upper tail-dependence coefficients. In Section 4.3 we show that the notion of a *tail copula* is justified because many properties of tail copulae are closely related to properties of copulae. In order to discuss asymptotic normality and strong consistency of the introduced estimators, we define a suited function space of locally uniformly bounded functions on compact sets in Section 4.4. Various results on asymptotic normality and strong consistency are elaborated in Sections 4.5 and 4.6 by means of a powerful functional Delta method stated in Van der Vaart and Wellner (1996). Therefore, the definition of weak convergence with respect to outer expectations turns out to be appropriate. The proof of the main result, namely Theorem 4.5.3, is accomplished in two steps. First we prove asymptotic normality for distributions with known margins (cf. Theorem 4.5.2) and in a second step we drop the assumption of known margins. To our knowledge, the latter results seem to be new and form extensions of the results in Huang (1992); however, the way of proof is different. Further, Section 4.7 considers a general rank order statistics of extreme events and establishes a corresponding weak convergence result. Similar considerations for usual rank order statistics can be found in Ruymgaart, Shorack, and van Zwet (1972), Ruymgaart (1974), Rüschendorf (1976), and Fermanian, Radulović, and Wegkamp (2002).

In Section 4.9, we concentrate on various estimation problems for elliptically contoured distributions. In particular, we propose an attractive parametric estimator for the tail-dependence coefficient which is based on formula (3.18), and suggest a robust estimator for the correlation coefficient matrix. Turning to arbitrary distributions, Section 4.10 provides a guideline with various estimators for the upper and lower tail-dependence coefficients, where we distinguish between parametric and nonparametric estimators as well as estimation methods based on the entire data set or based on extreme data only. We present various statistical and empirical properties of the estimators, and discuss their prospective fields of application. Further, in Section 4.11, an extensive simulation study compares the introduced estimators with respect to their finite sample characteristics, and illustrates some of the statistical properties established in the previous sections.

In Chapter 5 we discuss several models of multivariate integrative distributions by analyzing their dependence structure, parameter estimation, random vector generation, and applicability to fit financial asset-return data. Section 5.1 is devoted to several parametric distributions with generalized hyperbolic margins. In particular, multivariate generalized hyperbolic distributions are considered in Section 5.1.1 which, however, may lack of robust and fast estimation procedures within high-dimensional settings. Therefore, a new class of multivariate distributions is introduced in Section 5.1.2 belonging to affine-linear transformed random vectors with independent and generalized hyperbolic margins. These distributions possess good estimation properties, and have attractive dependence structures according to Section 5.1.3. Further, in Sections 5.1.5 and 5.1.6, we develop the necessary algorithms for parameter estimation and random number generation. In particular, we provide a pseudo software-code for two efficient random number generators which are based on a rejection algorithm with some simple envelop. The advantages and disadvantages of both types of distributions are discussed and illustrated via a simulation study in Sections 5.1.7 and 5.1.8.

Section 5.2 focuses on a family of semi-parametric multivariate distributions which primarily consists of a subclass of elliptically contoured distributions. In particular, we model the density generator of elliptical distributions by nonparametric methods, and the scaling matrix and the location vector by parametric methods. The multivariate (symmetric) generalized hyperbolic distributions form an important special case. The main emphasis is put on normal variance mixtures with self-decomposable mixing distributions. In Section 5.2.4, we fit the semi-parametric distributions to several financial time series, and discuss the Value-at-Risk for linear asset portfolios which are based on these distributions.

The last part of this thesis, Chapter 6, is devoted to special applications of the previously discussed distribution models and dependence measures. Section 6.1 provides an application of the latter models and techniques within the framework of a coherent risk-based pricing system. This system has been successfully implemented in cooperation with DaimlerChrysler AG Research and Technology in Ulm. Further, Section 6.2 exhibits the portfolio model which is utilized within the Internal Ratings based Approach of the New Basel Capital Accord in order to calculate regulatory banking capital. Section 6.3 shortly presents further outcomes of the above cooperation. Finally, in Section 6.4, the lower tail-dependence coefficient is estimated for several financial time series and some conclusions of the results are drawn.

Chapter 2

Basic concepts and definitions

2.1 Extreme events in financial markets

2.1.1 Historical data

During the last decade, severe financial crises of companies belonging either to old or new-economies in traditional or emerging markets have been often caused by a wrong or missing assessment of market risk and credit risk. Frequently, single extreme events like extremely large negative asset-returns (for example during a market crash or bear markets) account for most financial crises of banks or other companies. Some of these crises led directly to a default of the corresponding company or other investing institutions. The following list contains some well-known financial institutions (and the related losses): Barings ($1 billion), Allied Irish Bank ($700 millions), Daiwa ($1 billion), Kidder Peabody ($350 millions), LTCM ($4 billions), and Midland Bank ($500 millions). Besides, we list some non-financial institutions which have also been subject to large losses: Hammersmith and Fulham ($600 millions), Metallgesellschaft ($1.8 billions), Orange County ($2 billions), Procter and Gamble ($90 millions), and Shell ($1 billion). However, many financial crises of large companies contribute to a further destabilization of the corresponding industrial sector or of entire economies. This phenomenon has drawn the attention of regulators, financial analysts, portfolio risk-managers, and academic researchers to comoving micro- and macroeconomic fluctuations. In other words, the main question is whether large losses or large asset-value changes occur in an isolated manner or in a concurrent manner. The second case immediately leads to the conjecture of possible dependencies of extreme events. Concerning this problem, according to Hartmann, Straetmans, and de Vries (2001), the monograph of Morgenstern (1959) is the earliest systematical treatment of the phenomenon of financial markets which spill over to other economies or other countries. Morgenstern (1959) investigated the effects of 23 stock market crashes and explicitly refers to the "statistical extremes" of market movements. He observes significant dependencies between financial markets or financial assets, especially during stock market crashes. However, strong dependencies during stock market crashes can be already monitored without the application of sophisticated mathematical tools. Consider, for instance, the stock market crash after the events of September 11, 2001, or the crash on October 19, 1987, which is known as the "Black Monday". Regarding the latter crash, the Dow Jones dropped 508 points to 1738, the S&P 500 fell 58 points to 225, and the NASDAQ went down 46 points to 360 in only one day. Nearly all international stock markets followed the same trend, for example, the DAX fell 137 points to 1322 and the CAC40 dropped 129 points to 1209 in one day. Further, the NIKKEI 225 fell 4457 points to 21910 and the FTSE 100 dropped 500 points to 1802 in two days. A recent survey on contagion across international equity markets is given by Longin and Solnik (2001).

Two main factors causing changes in time of (extremal) asset dependencies have been discussed in the risk-management community: The first factor concerns the market volatility and the second factor relates to the market trend. As far as the first factor is concerned, several empirical and theoretical elaborations show evidence that dependencies between financial markets or asset returns increase with higher market volatility, and decrease with lower market volatility; references are King and Wadhwani (1990) and Ramchand and Susmel (1998). Many of these surveys utilize time-series models like G(ARCH)-type models to draw conclusions about varying dependencies (cf. Susmel and Engle (1994), Silvapulle and Granger (2001)). The second factor, although apparent, is not investigated that thoroughly in the literature; see, for example, Longin and Solnik (2001). Section 6.4 provides some empirical investigations regarding this topic.

2.1.2 Modelling extreme events

Extreme-value theory (EVT) plays a dominant role concerning inferences about extreme events and the tail behavior of the associated probability distributions (see, for example, de Haan (1994), Resnick (1987), Falk, Hüsler, and Reiss (1994), de Haan and Stadtmüller (1996)). In contrast to multivariate EVT, univariate EVT is well-established and applied in finance and insurance industry (cf. Embrechts, Klüppelberg, and Mikosch (1997)). The reasons for that are important theorems which have been mainly established in the one-dimensional setting: The Fisher-Tippett theorem, the Gnedenko theorem, and the Pickand-Balkema-de Haan theorem. For readers who are not familiar with EVT, we provide a short outline of these theorems.

The following definition is equivalent in the one-dimensional and multidimensional framework. Let $X,\ X^{(1)},\ X^{(2)},\dots,X^{(m)},\ m \in I\!N$, be independent, identically distributed n-dimensional random vectors with distribution function F. The primary concern of EVT relates to extreme values or maximal values (in practice: large asset returns or large portfolio losses) of the above random sample, for example, $\max_{1\leq j\leq m} X_i^{(j)},\ i = 1,\dots,n$. However, sensible assertions about extreme events can only be made if enough data to infer from are available. EVT provides techniques to trade off the bias of having insufficient data in practice and meaningful extrapolations beyond the range of given data. For this, the class of extreme value distributions turns out to be essential.

The random vector X or its distribution is said to be in the *domain of attraction* of a *multivariate extreme value distribution* G if there exists a sequence of normalizing constants $(a_{mi})_{m=1}^{\infty}$, $(b_{mi})_{m=1}^{\infty}$ with $a_{mi} > 0$ and $b_{mi} \in I\!R$, $i = 1, \ldots, n$, such that

$$\mathbb{P}\Big(\frac{\max_{1\le j\le m} X_1^{(j)} - b_{m1}}{a_{m1}} \le x_1,\ \ldots\ ,\frac{\max_{1\le j\le m} X_n^{(j)} - b_{mn}}{a_{mn}} \le x_n\Big)$$

converges to the value $G(x_1, \ldots, x_n)$ of the limit distribution function G with non-degenerate margins as $m \to \infty$ (The domain of G is specified below). In particular, the latter convergence is equivalent to

$$\lim_{m\to\infty} F^m(a_{m1}x_1 + b_{m1}, \ldots, a_{mn}x_n + b_{mn}) = G(x_1, \ldots, x_n). \tag{2.1}$$

In the one-dimensional setting, the Fisher-Tippett theorem provides a completely parametric characterization of the extreme value distributions G.

Theorem 2.1.1 (Fisher and L.H.C.Tippett (1928)) *Let the real-valued random variable X be in the domain of attraction of an extreme value distribution G. Then the corresponding distribution function $G : I\!R \to [0,1]$ must be of the form:*

$$G(x) = \begin{cases} \exp\Big\{-\Big(1+\xi\cdot\frac{x-\mu}{\sigma}\Big)^{-1/\xi}\Big\} & \text{if}\quad \xi \neq 0 \\ \exp\Big\{-\exp\Big(-\frac{x-\mu}{\sigma}\Big)\Big\} & \text{if}\quad \xi = 0 \end{cases} \tag{2.2}$$

for each $x \in I\!R$ such that $1 + \xi(x-\mu)/\sigma > 0$ and for some scaling parameter $\sigma > 0$, location parameter $\mu \in I\!R$, and shape parameter $\xi \in I\!R$.

Remark. The three different cases of the extreme value distribution G represented in (2.2) are called: Gumbel distribution if $\xi = 0$, Fréchet distribution if $\xi > 0$, and Weibull distribution if $\xi < 0$.

Fréchet distributions ($\xi > 0$) are primarily applied in practice because of their unbounded support on the positive halfline and their relationship to so-called heavy-tailed distributions. Another important property is specified by the well-known Gnedenko theorem.

Theorem 2.1.2 (Gnedenko (1943)) *The real-valued random variable X with distribution function F is in the domain of attraction of a Fréchet distribution ($\xi > 0$) if and only if its tail function $\bar{F} \equiv 1 - F$ is regularly varying (cf. Definition* 2.3.1*).*

The term "regularly varying tail function" in Theorem 2.1.2 is often referred to as heavy tailed distribution function in the literature. Returning to questions of risk management, the random variable X may represent certain portfolio losses or negative asset returns. It has been already mentioned that risk managers are typically concerned about the VaR related to X which is formally defined by

$$VaR = VaR^{\alpha} := -\text{arginf}_{y\in I\!R}\mathbb{P}(X \ge y) \ge 1-\alpha, \tag{2.3}$$

where the confidence level $\alpha > 0$ is assumed to be small. See also Section 5.2.4 for a detailed treatment. Usually the absolute value of VaR corresponds to the amount of economic capital or regulatory capital a financial institution needs to set aside in order to cover large losses in its portfolio. However, there still is a small probability α that the institution suffers a larger loss and a possible default. Therefore, risk managers are also interested in the conditional distribution function of excesses G_{VaR}, i.e.,

$$G_{VaR}(x) = \mathbb{P}(X + VaR \leq x \mid X > -VaR). \tag{2.4}$$

The Pickands-Balkema-de Haan theorem connects the distribution function G_{VaR} with the already introduced techniques from EVT.

Theorem 2.1.3 (Pickands (1975), Balkema and de Haan (1974))
Let the random variable X have the distribution function F with right endpoint x_F. Then X is in the domain of attraction of an extreme value distribution G if and only if the following uniform convergence holds

$$\lim_{VaR \uparrow x_F} \sup_{0 < x - \mu < x_F - VaR} \Big| G_{VaR}(x) - (1 + \log G(x)) \Big| = 0,$$

where the distributions function G_{VaR} is defined in (2.4) and the scale parameter σ of the extreme value distribution G is a function of the VaR, i.e., $\sigma(VaR)$.

Summarizing the above results, univariate EVT provides probabilistic tools to model the limiting distributions of normalized maxima and excesses over high thresholds. Regarding the parameter estimation of extreme value distributions it suffices to apply parametric estimation methods instead of nonparametric estimation methods which are less robust for small sample sizes. Multivariate extreme-value theory turns out to be more complicated, as in general no analogue of the above Fisher-Tippett theorem exists. Instead multivariate extreme value distributions are characterized by a parametric and a nonparametric component.

Suppose the n-dimensional random vector X is in the domain of attraction of an extreme value distribution G. A necessary condition is obvious, namely all one-dimensional margins of X must be in the domain of attraction of some extreme value distribution. Moreover, the one-dimensional convergence takes place with the same normalizing constants; i.e., for the bivariate case we have (here F_1, F_2 and G_1, G_2 denote the marginal distribution functions):

$$\begin{aligned} \lim_{m \to \infty} m(1 - F_1(a_{m1}x_1 + b_{m1})) &= -\log G_1(x_1) \quad \text{and} \\ \lim_{m \to \infty} m(1 - F_2(a_{m2}x_2 + b_{m2})) &= -\log G_2(x_2) \end{aligned} \tag{2.5}$$

which is equivalent to the convergence (2.1) in the one-dimensional setting. However, the convergence in (2.5) is also equivalent to

$$\begin{aligned} \lim_{m \to \infty} \frac{(1 - F_1)^{-1}(x_1/m) - b_{m1}}{a_{m1}} &= (-\log G_1)^{-1}(x_1) \quad \text{and} \\ \lim_{m \to \infty} \frac{(1 - F_2)^{-1}(x_2/m) - b_{m2}}{a_{m2}} &= (-\log G_2)^{-1}(x_2) \end{aligned} \tag{2.6}$$

for $x_1,\ x_2 > 0$ and H^{-1} denoting the generalized inverse function of the monotone function $H : D \subset I\!R \to I\!R$, i.e.,

$$H^{-1}(x) := \inf\{y \in I\!R \mid H(y) \geq x\}, \quad x \in D \subset I\!R, \tag{2.7}$$

and $H^{-1}(x) := \infty$ in case $\inf\{y \in I\!R \mid H(y) \geq x\} = \emptyset$. This observation, together with (2.1), leads to the existence of the following limit

$$\lim_{m\to\infty} m(1 - F((1-F_1)^{-1}(x_1/m), (1-F_2)^{-1}(x_2/m))) =: l(x_1, x_2), \quad x_1, x_2 > 0, \tag{2.8}$$

where $l(x_1, x_2)$ is referred to as the so-called *stable tail-dependence function* (cf. Huang (1992), p. 26). In particular, the relationship between a bivariate extreme value distribution G and the corresponding stable tail-dependence function l is given by:

$$l(x_1, x_2) = -\log G((-\log G_1)^{-1}(x_1), (-\log G_2)^{-1}(x_2)), \quad x_1, x_2 > 0. \tag{2.9}$$

On the other hand a sufficient condition for an n-dimensional random vector X or its distribution function F to be in the domain of attraction of a multivariate extreme value distribution is that the corresponding copula of F and all one-dimensional margins are in the domain of attraction of an extreme value distribution. Denote by $\mathbb{S}^{n-1} := \{x \in I\!R^n : ||x|| = 1\}$ the $(n-1)$-dimensional unit sphere for some arbitrary norm $||\cdot||$ in $I\!R^n$. After a specific standardization (see Section 3.3.2), the standardized distribution function G_* can be represented as

$$G_*(x) = \exp\Big(-\int_{\overline{\mathbb{S}}^{n-1}} \big[\max\big\{\frac{a_i}{x_i}, i \in I_a\big\}\big]^\alpha\ S(da)\Big), \quad x \in I\!R^n_+,$$

where $S(\cdot)$ denotes the so-called spectral measure living on the n-dimensional unit sphere $\mathbb{S}^{n-1}$, $\overline{\mathbb{S}}^{n-1} := \mathbb{S}^{n-1} \backslash (-\infty, 0]^n$, and $I_a = \{j \in \{1, \ldots, n\} \mid a_j > 0\}$. A multivariate version of the Gnedenko theorem can be also formulated; more details will be elaborated in Section 3.3.2. References on multidimensional extreme-value theory are Resnick (1987), Chapter 5, Tawn (1988), de Haan and Resnick (1993), and Huang (1992).

2.1.3 How to observe extremal dependence?

Regarding the discussion in Section 2.1.1 and the paragraph before that section, the following question arises: Is it actually possible to detect or measure dependencies of extreme events in finance by means of statistical tools? The present section provides several exploratory visual aids to detect extremal dependencies in financial time series. The scatter plots in Figure 2.1 serve as a first clue that extremal dependencies might indeed be found in many financial data.

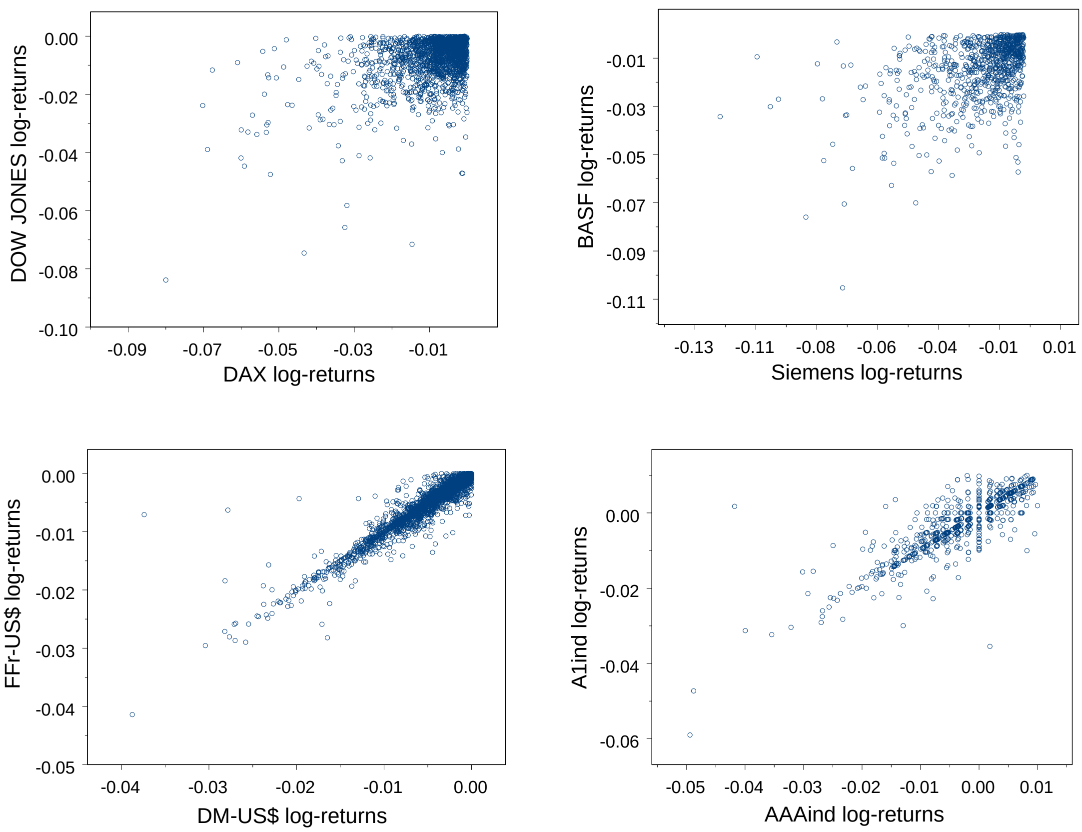


Figure 2.1: Scatter plots of DAX versus DOW JONES (upper left plot) daily index log-returns (7255 data points, time frame: 1975-2002), Siemens versus BASF (upper right plot) daily stock log-returns (2699 data points, time frame: 1992-2002), FFr-US$ versus DM-US$ (lower left plot) daily exchange-rate log-returns (4634 data points, time frame: 1984-2002), and AAAind versus A1ind (lower right plot) daily index log-returns (1391 data points, time frame: 1995-2000). The extreme log-return on 19th of October 1987 "Black Monday" is omitted in the first data-set for reasons of visualization.

For the remaining of this section we concentrate on two pairs of time-series, namely the log-returns of DAX-DOW JONES stock indices and the log-returns of Siemens-BASF stocks. Both time series can be assumed to be bivariate elliptically contoured distributed (see Section 3.2 for a definition of elliptically contoured distributions and Section 5.2.4 for tests on elliptical symmetry). In Section 3.2 we will show that the existence of extremal dependence for elliptically contoured distributions is nearly equivalent with the existence of heavy-tailed one-dimensional marginal distribution functions. Therefore, we can also utilize techniques from univariate EVT (cf. Embrechts, Klüppelberg, and Mikosch (1997)) to detect heavy-tailed margins; standard visualization techniques are mean excess plots or QQ-plots. The QQ-plots in Figure 2.2 compare the quantiles of the empirical distribution of the latter time-series with quantiles of the normal distribution (which is known to be not heavy-tailed). The plots clearly reveal that all time-series posses much heavier tails.

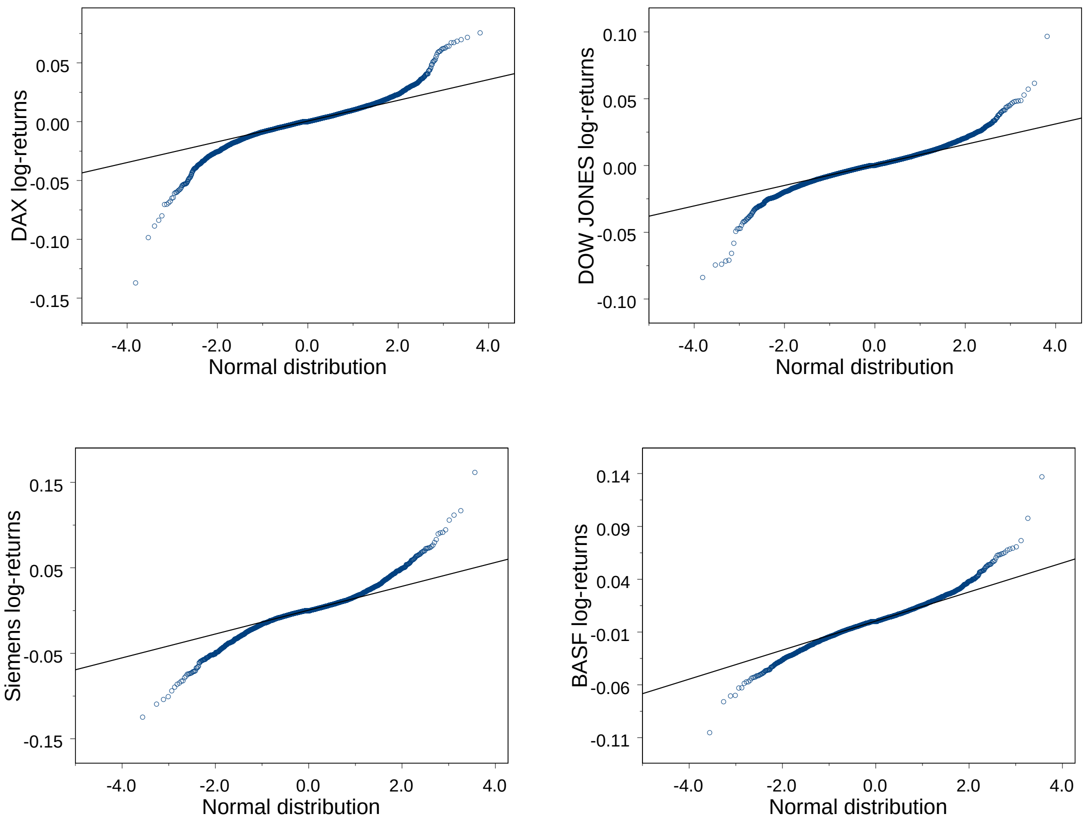


Figure 2.2: QQ-plots of DAX, DOW JONES, Siemens, and BASF daily index or stock log-returns versus normal distribution. The extreme log-return on 19th of October 1987 "Black Monday" is omitted in the DOW JONES QQ-plot for reasons of visualization.

Recall that in most financial applications the correlation coefficient still represents the only dependence measure between two random variables. For reasons of comparison we provide in Figure 2.3 the plots of the empirical correlation coefficients over time which correspond to the above time-series.

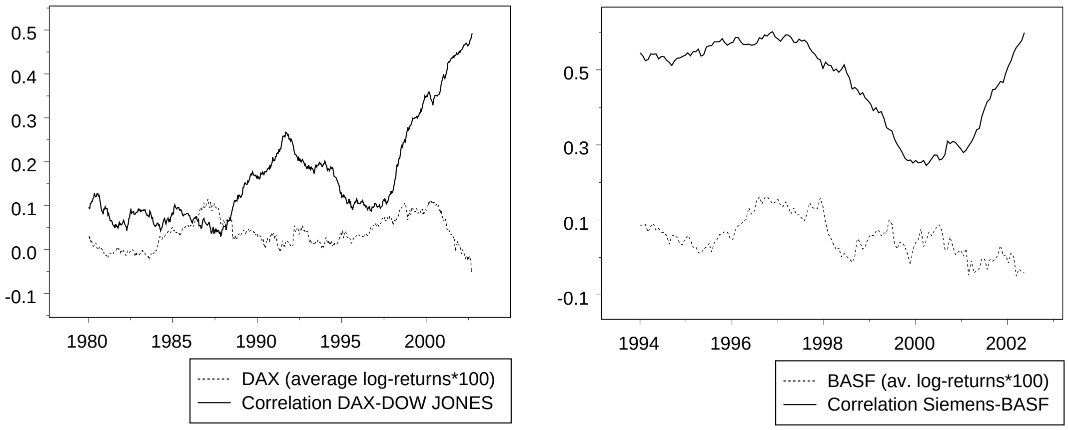


Figure 2.3: Plots of empirical correlation coefficients over time between DAX and DOW JONES (left plot) daily index log-returns (correlations are calculated for the years 1975-2002 with a horizon of 1000 trading days and a successive delay of 11 trading days) and between Siemens and BASF (right plot) daily stock log-returns (correlations are calculated for the years 1992-2002 with a horizon of 500 trading days and a successive delay of 15 trading days).

However, as we have already mentioned, the correlation coefficient is often a non-appropriate dependence measure when leaving the Gaussian world. Instead, the copula concept turns out to be the right concept when we talk about dependencies of multivariate distributions in general. Figure 2.4 presents the empirical copula density of the above time-series. The clustering in the upper right and lower left corner of the empirical copula density is a strong evidence for extremal dependence.

Consequently, practitioners should primarily consider dependence measures which dependent only on the copula of the underlying random vector. This is unfortunately not true for the correlation coefficient which depends strongly on the marginal distributions, especially if we are not in the framework of elliptically contoured distributions. However, Kendall's tau (cf. Definition 5.1.7) represents an attractive alternative because it depends only on the corresponding copula and it is as easy to calculate as the correlation coefficient. Note that the behavior of empirical Kendall's tau in Figure 2.5 is quite similar to the behavior of the empirical correlation coefficients in Figure 2.3.

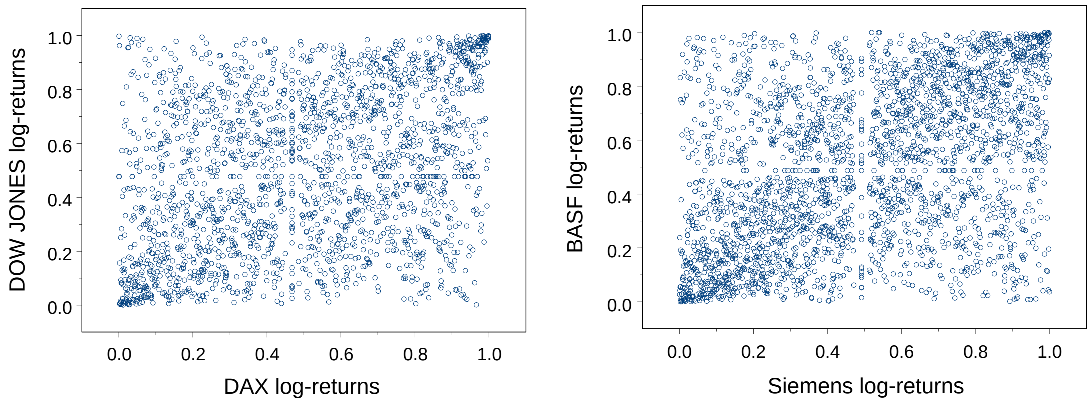


Figure 2.4: Empirical copula density plots of DAX versus DOW JONES (left plot) daily index log-returns (2037 data points, time frame: 1995-2002) and Siemens versus BASF (right plot) daily stock log-returns (2699 data points, time frame: 1992-2002).

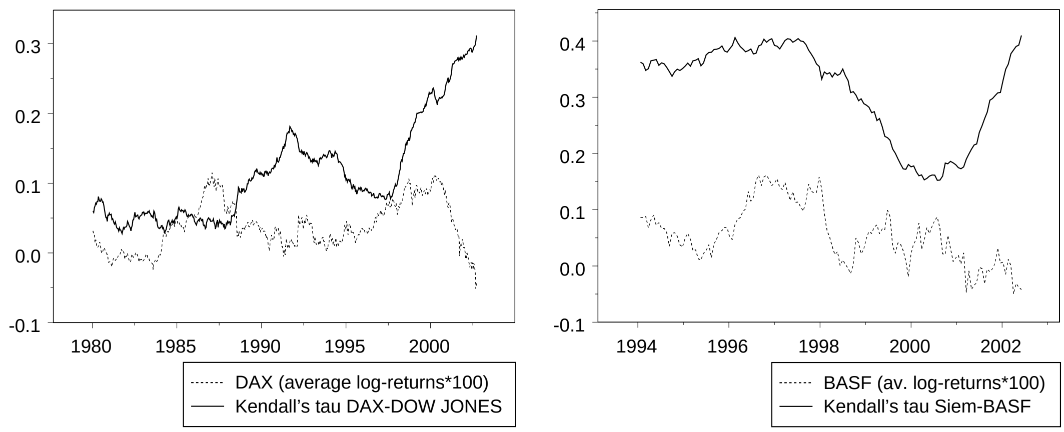


Figure 2.5: Plots of empirical Kendall's tau over time between DAX and DOW JONES (left plot) daily index log-returns (Kendall's tau are calculated for the years 1975-2002 with a horizon of 1000 trading days and a successive delay of 11 trading days) and between Siemens and BASF (right plot) daily stock log-returns (Kendall's tau are calculated for the years 1992-2002 with a horizon of 500 trading days and a successive delay of 15 trading days).

The main aim of this thesis is to provide more inside into the fields of dependence structures and dependence measures, especially in the context of extreme events. Regarding extremal dependence, the (upper and lower) tail-dependence coefficient turns out to be a quite useful tool; not only because it depends solely on the copula. Although we have not defined the (upper and lower) tail-dependence coefficient yet, Figure 2.6 represents the (lower) tail-dependence coefficient estimates over time for the above time series. The main message is that extremal dependence may increase dramatically during bear markets or market crashes, and the behavior of the tail-dependence coefficient might be quite different from other dependence measures, like the correlation coefficient or Kendall's tau.

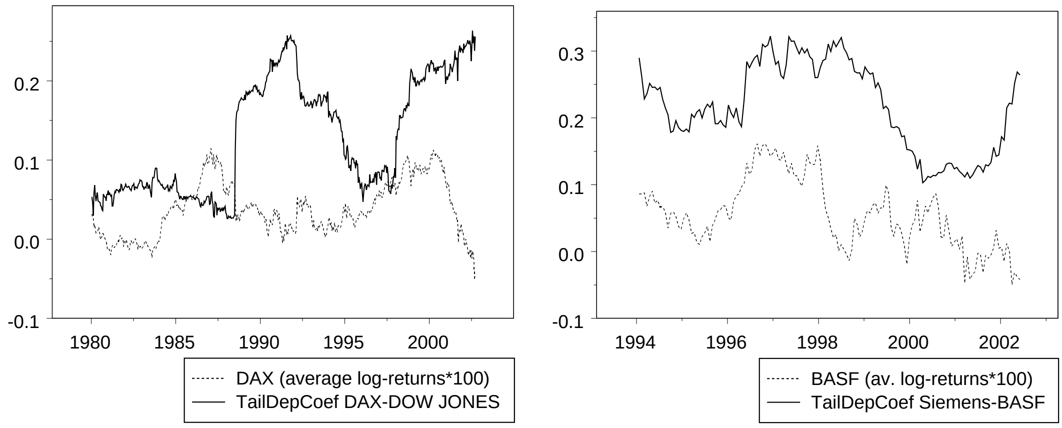


Figure 2.6: Plots of the estimates of the (lower) tail-dependence coefficient over time between DAX and DOW JONES (left plot) daily index log-returns (TDCs are calculated for the years 1975-2002 with a horizon of 1000 trading days and a successive delay of 11 trading days) and between Siemens and BASF (right plot) daily stock log-returns (TDCs are calculated for the years 1992-2002 with a horizon of 500 trading days and a successive delay of 15 trading days).

2.2 Dependence modelling

2.2.1 Copulae: Definitions and basic properties

The theory of copulae deals with the dependence structure of multidimensional random vectors. On the one hand, copulae are functions that join or "couple" multivariate distribution functions to their corresponding marginal distribution functions. On the other hand, a copula function itself is a multivariate distribution function with uniform margins on the interval $[0, 1]$. Copulae are of interest in finance because of two reasons: First, as a way of studying the dependence structure of an asset portfolio irrespective of its marginal asset-return distributions; and second, as a starting point for constructing multidimensional distributions for asset portfolios, with a view to simulation. First we define the copula function in a common way (see Joe (1997), p. 12).

Definition 2.2.1 *Let $C : [0,1]^n \to [0,1]$ be an n-dimensional distribution function on $[0,1]^n$. Then C is called a copula if it has uniformly distributed one-dimensional margins on the interval* $[0,1]$.

The following theorem gives the foundation for a copula to inherit the dependence structure of a multidimensional distribution.

Theorem 2.2.2 (Sklar's theorem) *Let F be an n-dimensional distribution function with margins $F_1, \dots, F_n$. Then there exists a copula C, such that for all $x \in \mathbb{R}^n$*

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)). \tag{2.10}$$

*If $F_1, \dots, F_n$ are all continuous, then C is unique; otherwise C is uniquely determined on $RanF_1 \times \cdots \times RanF_n$, where Ran refers to the corresponding range. Conversely, if C is a copula and $F_1, \dots, F_n$ are one-dimensional distribution functions, then the function F given by (*2.10*) is an n-dimensional distribution function with margins $F_1, \dots, F_n$.*

For a *proof* we refer the reader to Sklar (1996) or Nelsen (1999).

The following corollary is immediately obtained from Theorem 2.2.2. It shows how one can construct the copula of a multidimensional distribution function.

Corollary 2.2.3 *Let F be an n-dimensional continuous distribution function with margins $F_1, \dots, F_n$. Then the corresponding copula C has representation*

$$C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)), \quad 0 \le u_1, \dots, u_n \le 1,$$

where $F_1^{-1}, \dots, F_n^{-1}$ denote the generalized inverse (distribution) functions of $F_1, \dots, F_n$ defined by (2.7).

According to Schweizer and Wolff (1981): "... the copula is invariant while the margins may be changed at will, it follows that it is precisely the copula which captures those properties of the joint distribution which are invariant under a.s. strictly increasing transformations" and thus the copula function represents the dependence structure of a multivariate random vector. We add some more copula properties needed later.

Remarks.

1. If an n-dimensional copula-density c exists, then a density f of F exists and the following relationship holds:

$$f(x_1, \dots, x_n) = c(F_1(x_1), \dots, F_n(x_n)) \times \prod_{i=1}^{n} f_n(x_n), \tag{2.11}$$

   where F_i and f_i denote the corresponding marginal distributions and density functions, respectively.

2. A copula is increasing in each component. In particular the partial derivatives $\partial C(u)/\partial u_i,\ i = 1 \dots n$, exist almost everywhere.

3. Consequently, the conditional distributions of the form

$$C(u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_n \mid u_j), \quad j = 1, \dots, n, \tag{2.12}$$

exist.

4. A copula C is uniformly continuous on $[0, 1]^n$.

For more details regarding the theory of copulae we refer the reader to the monographs of Nelsen (1999) and Joe (1997).

2.2.2 Measures of extremal dependence

Regarding the hypothesis of increasing dependencies in financial markets, most academic investigations utilize traditional dependence measures like covariances or correlations or, turning to dependence measures of extreme events, conditional correlations (cf. Loretan (2000), Malevergne and Sornette (2002)). However, these dependence measures may heavily depend on the shape of the marginal distribution functions, which can distort the real dependence structure and consequently may yield wrong implications (cf. Embrechts, McNeil, and Straumann (1999)). Recall, it is exactly the copula which separates the dependence structure of multivariate distribution functions from their marginal distribution functions. Therefore, we rather prefer dependence measures which are solely based on the copula of the corresponding multivariate distribution, i.e., "scale-invariant" dependence measures. One potential candidate, investigated in this thesis, might be the tail-dependence coefficient defined in Section 2.2.3 below which relates to dependencies of extreme events. There is hardly any literature about the latter type of dependence measures and this thesis aims to fill that gap. Some references are Hauksson et al. (2001), Malevergne and Sornette (2002), and Lindskog (1999).

2.2.3 Tail dependence: A copula property

In this section we formally introduce the concept of tail dependence and embed it into the framework of copulae. In the literature, the definition of tail-dependence for multivariate distributions is mostly related to their bivariate marginal distributions. They reflect the limiting proportion of exceedance of one margin over a certain threshold given that the other margin has already exceeded that threshold. The following approach represents one of many possible definitions of tail dependence.

Definition 2.2.4 (Tail dependence, (Joe 1997), p. 33) *Let $X = (X_1, X_2)'$ be a 2-dimensional random vector. We say that X is upper tail-dependent if*

$$\lambda_U := \lim_{u \to 1^-} \mathbb{P}(X_1 > F_1^{-1}(u) \mid X_2 > F_2^{-1}(u)) > 0, \tag{2.13}$$

where the limit is assumed to exist and F_1^{-1}, F_2^{-1} *denote the generalized inverse distribution functions of* X_1, X_2, *respectively. Consequently, we say that* X *is* upper tail-independent *if* λ_U *equals* 0. *Further,* λ_U *is called the* upper tail-dependence coefficient *(upper TDC). Similarly, the* lower tail-dependence coefficient *(lower TDC) is defined by*

$$\lambda_L := \lim_{u \to 0^+} \mathbb{P}(X_1 \le F_1^{-1}(u) \mid X_2 \le F_2^{-1}(u))$$

if existent and X *is said to be* lower tail-dependent *(*lower tail-independent*) if* $\lambda_L > 0$ $(\lambda_L = 0)$.

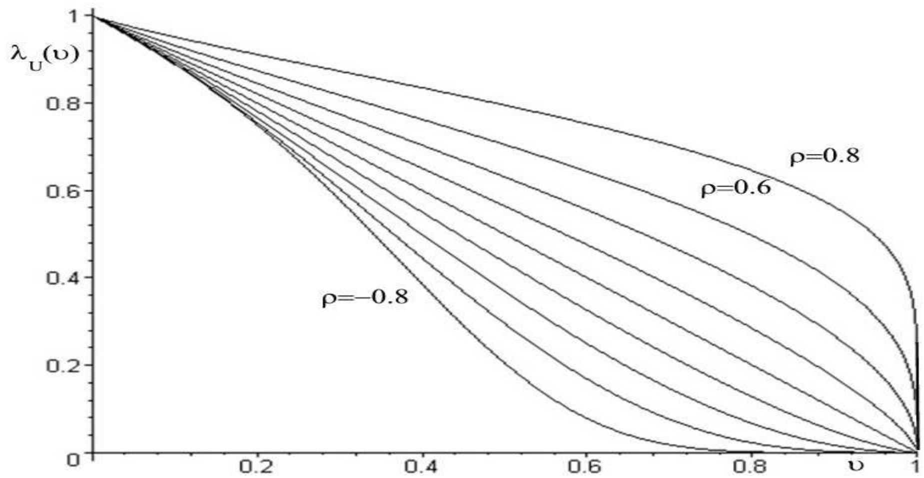


Figure 2.7: $\lambda_U(u) = \mathbb{P}(X_1 > F_1^{-1}(u) \mid X_2 > F_2^{-1}(u))$ for bivariate normal distributions with $\rho = -0.8,\ -0.6, \ldots, 0.6,\ 0.8$. In this case $\lambda_U = 0$ for all $\rho \in (-1, 1)$.

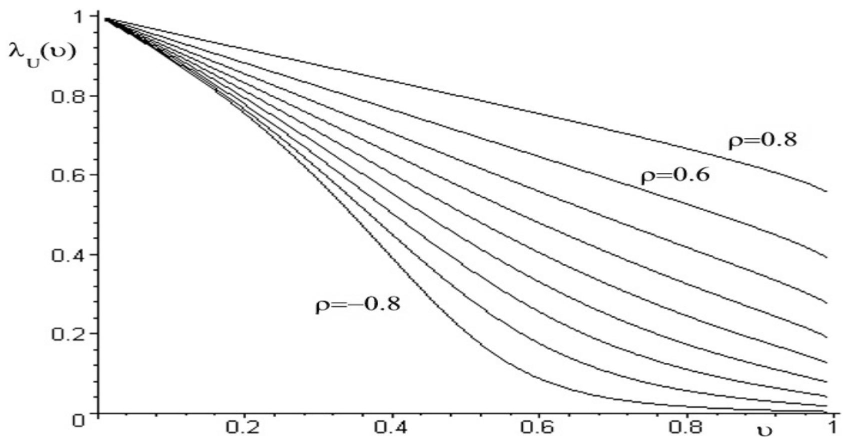


Figure 2.8: $\lambda_U(u) = \mathbb{P}(X_1 > F_1^{-1}(u) \mid X_2 > F_2^{-1}(u))$ for bivariate t-distributions with $\rho = -0.8,\ -0.6, \ldots, 0.6,\ 0.8$. In this case $\lambda_U > 0$ for all $\rho \in (0, 1)$. The exact formula for λ_U is given in Section 3.2.

The following proposition shows that tail dependence is a copula property. Thus many copula features can be passed to the tail-dependence coefficient, for example the invariance under strictly increasing transformations of the margins.

Proposition 2.2.5 *Let X be a continuous bivariate random vector, then*

$$\lambda_U = \lim_{u \to 1^-} \frac{1 - 2u + C(u,u)}{1-u} \tag{2.14}$$

provided that the limit exists, where C denotes the copula of X. Analogously, $\lambda_L = \lim_{u\to 0^+} \frac{C(u,u)}{u}$ holds for the lower tail-dependence coefficient.

Proof. Let F_1 and F_2 be the marginal distribution functions of X. Then

$$\begin{aligned}
\lambda_U &= \lim_{u\to 1^-} \mathbb{P}(X_1 > F_1^{-1}(u) \mid X_2 > F_2^{-1}(u)) \\
&= \lim_{u\to 1^-} \frac{\mathbb{P}(X_1 > F_1^{-1}(u), X_2 > F_2^{-1}(u))}{\mathbb{P}(X_2 > F_2^{-1}(u))} \\
&= \lim_{u\to 1^-} \frac{1 - F_2(F_2^{-1}(u)) - F_1(F_1^{-1}(u)) + C(F_1(F_1^{-1}(u)), F_2(F_2^{-1}(u)))}{1 - F_2(F_2^{-1}(u))} \\
&= \lim_{u\to 1^-} \frac{1-2u+C(u,u)}{1-u}.
\end{aligned}$$

□

Notice that according to Proposition 2.2.5, the upper and lower tail-dependence coefficient are solely determined by the diagonal section of the corresponding copula. The clustering in the upper right and lower left corner of the empirical copula densities in Figure 2.4 justifies the restriction to this copula section. In particular, if we are interested in a single figure describing the amount of extremal dependence of a bivariate distribution or copula, the tail-dependence coefficients seem to be the right choice.

Although we provided a simple characterization for upper and lower tail dependence by Proposition 2.2.5, it will be still difficult and tedious to verify tail dependence or tail independence if the copula is not given via a closed-form expression. This is the case, for example, for many well-known elliptically contoured distributions. Therefore, the following result provides another approach for computing the tail-dependence coefficients. We restrict ourselves to upper tail dependence.

Proposition 2.2.6 *Let X be a bivariate random vector with copula C which is totally differentiable on $(0,1)^2$. Then the upper tail-dependence coefficient λ_U can be expressed using conditional probabilities provided that the following limit exists:*

$$\lambda_U = \lim_{u\to 1^-} \big(\mathbb{P}(U_1 > u \mid U_2 = u) + \mathbb{P}(U_2 > u \mid U_1 = u)\big), \tag{2.15}$$

where the random vector (U_1, U_2) is distributed according to the copula C of X.

Proof. According to Theorem 7.29 in Wheeden and Zygmund (1977), the conditional probability functions in (2.15) exist because copula functions $C(u,v)$ are Lipschitz-continuous (cf. Theorem 2.2.4 in Nelsen (1999)). Further, we may write $\mathbb{P}(U_1 \leq u \mid U_2 = v) = \partial C(u,v)/\partial v$ and $\mathbb{P}(U_1 > u \mid U_2 = v) = 1 - \partial C(u,v)/\partial v$, respectively. Then, the rule of l'Hospital and Theorem 166.1 in Heuser (2000) implies that

$$\begin{aligned}
\lambda_U &= \lim_{u \to 1^-} \frac{1 - 2u + C(u,u)}{1-u} = \lim_{u \to 1^-} \Big(- \Big(-2 + \frac{dC(u,u)}{du} \Big) \Big) \\
&= \lim_{u \to 1^-} \Big(2 - \frac{\partial C(x,u)}{\partial x}\Big|_{x=u} - \frac{\partial C(u,y)}{\partial y}\Big|_{y=u} \Big) \\
&= \lim_{u \to 1^-} \big(\mathbb{P}(U_1 > u \mid U_2 = u) + \mathbb{P}(U_2 > u \mid U_1 = u) \big).
\end{aligned}$$

□

2.2.4 Tail copulae and tail dependence

Tail copulae provide a more general concept for the dependence modelling of extreme events. The properties of tail copulae are closely related to those of copulae; cf. Section 4.3. This fact justifies the notion *tail copula*. The following definition of a tail copula is formulated such that it represents an extension of the concept of tail dependence discussed in Section 2.2.3 (cf. Definition 2.2.4). The relationship to other concepts of dependence modelling for extreme events is given below.

Definition 2.2.7 (Tail copulae) *Define $\bar{\mathbb{R}}^n_+ := [0,\infty]^n \backslash \{(\infty,\ldots,\infty)\}$. The function $\Lambda_U^{I,J} : \bar{\mathbb{R}}^n_+ \to \mathbb{R}$, $I, J \subset \{1,\ldots,n\}$, $I \cap J = \emptyset$ is called an upper tail copula associated with the n-dimensional distribution function F if the following limit exists everywhere*

$$\Lambda_U^{I,J}(x) := \lim_{t\to\infty} \mathbb{P}(X_i > F_i^{-1}(1 - x_i/t),\ \forall i \in I \mid X_j > F_j^{-1}(1 - x_j/t),\ \forall j \in J). \quad (2.16)$$

The corresponding lower tail copula is defined by

$$\Lambda_L^{I,J}(x) := \lim_{t\to\infty} \mathbb{P}(X_i \leq F_i^{-1}(x_i/t),\ \forall i \in I \mid X_j \leq F_j^{-1}(x_j/t),\ \forall j \in J) \quad (2.17)$$

provided the limit exists.

In the two-dimensional setting we will consider a slight modification of the above defined tail copula, namely:

$$\begin{aligned}
\Lambda_U(x_1,x_2) &:= x_2 \cdot \Lambda_U^{\{1\},\{2\}}(x_1,x_2) \quad \text{and} && (2.18) \\
\Lambda_L(x_1,x_2) &:= x_2 \cdot \Lambda_L^{\{1\},\{2\}}(x_1,x_2), \quad x_1 \in \bar{\mathbb{R}}_+, x_2 \in \mathbb{R}_+, && (2.19)
\end{aligned}$$

where the indices $\{1\}$ and $\{2\}$ can be dropped in this case. Further, set $\Lambda_U(x_1,\infty) := x_1$ and $\Lambda_L(x_1,\infty) := x_1$ for all $x_1 \in \mathbb{R}_+$.

It is important to emphasize the relationship between the following notions: Tail copula, stable tail-dependence function, extreme value distribution and extreme value copula. According to Definition 2.2.7, the existence of a bivariate upper tail copula Λ_U is equivalent to the existence of a stable tail-dependence function l, introduced in (2.8). In particular, both functions are related to each other by

$$\Lambda_U(x_1, x_2) = x_1 + x_2 - l(x_1, x_2), \quad x_1, x_2 > 0.$$

Moreover, the tail copula and the stable tail-dependence function of a bivariate random vector X or its distribution function F exist if the corresponding copula function is in the domain of attraction of an extreme value distribution. It is important to notice that the distribution function F itself does not need to be in the domain of attraction of some extreme value distribution. This fact is crucial for practical applications where marginal distributions are often modelled separately from the copula function. However, if the distribution function F is in the domain of attraction of an extreme value distribution G, then the relationship between the upper tail copula Λ_U, the extreme value distribution G, and the corresponding extreme value copula C_G is given by

$$\begin{aligned} \Lambda_U(x_1, x_2) &= x_1 + x_2 + \log G((-\log G_1)^{-1}(x_1), (-\log G_2)^{-1}(x_2)) \\ &= x_1 + x_2 + \log C_G(\exp(-x_1), \exp(-x_2)), \quad x_1, x_2 > 0. \end{aligned} \tag{2.20}$$

References on topics related to extreme value copulae are Pickands (1981), Tawn (1988), and Cuculescu and Theodorescu (2002).

At this place we restate the definition of tail dependence and embed it into the concept of tail copulae.

Definition 2.2.4 *The bivariate random vector $(X_1, X_2)'$ is said to be upper tail-dependent if $\Lambda_U(1,1)$ exists and*

$$\lambda_U := \Lambda_U(1,1) = \lim_{v \to 1^-} \mathbb{P}(X_1 > F_1^{-1}(v) \mid X_2 > F_2^{-1}(v)) > 0. \tag{2.21}$$

Consequently, $(X_1, X_2)'$ is called upper tail-independent if λ_U equals 0. Similarly, the lower tail-dependence coefficient is defined by

$$\lambda_L := \Lambda_L(1,1) \tag{2.22}$$

if existent and lower tail-dependence (independence) is present if $\lambda_L > 0$ ($\lambda_L = 0$).

Recall that tail dependence is solely determined by the diagonal section of the corresponding copula (see Cuculescu and Theodorescu (2001) for an interesting account on diagonal sections). Therefore, for example, the copula function is not necessarily required to be in the domain of attraction of an extreme value copula. This property might be important for practical applications.

2.3 Further notions and results

2.3.1 Regular and O-regular variation

Some of the following results for copulae of elliptically contoured distributions are characterized by regularly varying or O-regularly varying functions and multivariate regularly varying random vectors, which are defined as follows.

Definition 2.3.1 *(Regular and O-regular variation)*

1. *The measurable function* $f : \mathbb{R}_+ \to \mathbb{R}_+$ *is called regularly varying (at* ∞*) with index* $\alpha \in \mathbb{R}$ *if for any* $t > 0$

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^\alpha.$$

   The class of regularly varying functions with index α *is denoted by* RV_α*.*

2. *The measurable function* $f : \mathbb{R}_+ \to \mathbb{R}_+$ *is called O-regularly varying (at* ∞*) if for any* $t \geq 1$

$$0 < \liminf_{x\to\infty} \frac{f(tx)}{f(x)} \leq \limsup_{x\to\infty} \frac{f(tx)}{f(x)} < \infty.$$

   The class of O-regularly varying functions is denoted by OR*.*

Thus, regularly varying functions behave asymptotically like power functions. Next we state a well-known uniform convergence theorem (see Resnick (1987), p.22) which will be frequently used in this thesis.

Theorem 2.3.2 *Let* $f \in RV_\alpha,\ \alpha \in \mathbb{R}$*.*

1. *Then*

$$\lim_{x\to\infty} \frac{f(xt)}{f(x)} = t^\alpha$$

   holds locally uniformly on $(0, \infty)$*. Moreover, if* $\alpha < 0$*, then uniform convergence holds on intervals of the form* $[b, \infty),\ b > 0$*.*

2. *Let* $\varepsilon > 0$*. Then there exists* $K_\varepsilon \geq 0$ *such that*

$$(1-\varepsilon)t^{\alpha-\varepsilon} < \frac{f(xt)}{f(x)} < (1+\varepsilon)t^{\alpha+\varepsilon}$$

   for all $x \geq K_\varepsilon$ *and* $t \geq 1$*.*

Further, we need the following result on the derivative of regularly varying functions. For a *proof* we refer to Bingham, Goldie, and Teugels (1987).

Theorem 2.3.3 *Suppose that the distribution function $F : \mathbb{R}_+ \to [0,1]$ is absolutely continuous with density f, i.e.,*

$$F(x) = \int_0^x f(t)\, dt, \qquad x \geq 0.$$

If $\overline{F} \equiv 1 - F \in RV_{-\alpha}$, $\alpha > 0$, and f is monotone, then $f \in RV_{-\alpha-1}$.

We shall also consider a Wiener-Tauberian theorem dealing with Mellin transform and Mellin convolution which we define now.

Definition 2.3.4 (Mellin transform and convolution)

1. *Given a measurable kernel $k : (0,\infty) \to \mathbb{R}$ we define its Mellin transform by*

$$\check{k}(z) := \int_0^\infty t^{-z} k(t) \frac{dt}{t} \tag{2.23}$$

   for $z \in \mathbf{C}$ such that the integral converges.

2. *For measurable functions $f, g : (0,\infty) \to \mathbb{R}$ we call $g \overset{M}{*} f$ the Mellin convolution, where*

$$g \overset{M}{*} f(x) := \int_0^\infty f(x/t) g(t) \frac{dt}{t} \tag{2.24}$$

   for those $x > 0$ for which the integral converges.

2.3.2 Multivariate regular variation

Definition 2.3.5 (Multivariate regular variation of random vectors) *The n-dimensional random vector $X = (X_1, \ldots, X_n)'$ and its distribution are said to be regularly varying with limit measure ν if there exists a function $b(t) \nearrow \infty$ as $t \to \infty$ and a non-negative Radon measure $\nu \neq 0$ such that*

$$t\mathbb{P}\Big(\Big(\frac{X_1}{b(t)}, \ldots, \frac{X_n}{b(t)}\Big) \in \cdot\Big) \xrightarrow{v} \nu(\cdot) \tag{2.25}$$

on the space $E = [-\infty,\infty]^n \backslash \{0\}$.

Notice that convergence $\xrightarrow{v}$ stands for vague convergence of measures, in the sense of Resnick (1987), p. 140. It can be shown that (2.25) requires the existence of a constant $\alpha \geq 0$ such that for each relatively compact set $B \subset E$ (i.e., the closure $\overline{B}$ of B is compact in E)

$$\nu(tB) = t^{-\alpha}\nu(B), \quad t > 0. \tag{2.26}$$

Thus we say that X is regularly varying with limit measure ν and index $-\alpha \leq 0$ if (2.25) holds. Moreover, the function $b(\cdot)$ is necessarily regular varying with index $-1/\alpha$. In the following we always assume that $\alpha > 0$.

Later, when we consider elliptically contoured distributions, it turns out that the use of polar coordinates is a convenient way to deal with multivariate regular variation. Recall that $\mathbb{S}^{n-1} = \{x \in I\!R^n : ||x|| = 1\}$ denotes the $(n-1)$-dimensional unit sphere for some arbitrary norm $||\cdot||$ in $I\!R^n$. Then the transformation into polar coordinates $T : I\!R^n \backslash \{0\} \to (0, \infty) \times \mathbb{S}^{n-1}$ is defined by

$$T(x) = \big(||x||, \frac{x}{||x||}\big) =: (r, a).$$

Observe that the point x can be seen as having distance r from the origin $0 = (0, \ldots, 0)'$ and direction $a \in \mathbb{S}^{n-1}$. It is well known that T is a bijection with inverse transform $T^{-1} : (0, \infty) \times \mathbb{S}^{n-1} \to I\!R^n \backslash \{0\}$ given by $T^{-1}(r, a) = ra$. For notational convenience we denote the Euclidian norm by $||\cdot||_2$ and the related unit sphere by $\mathbb{S}_2^{n-1}$. The next proposition, essentially stated in Resnick (2002), Proposition 2, characterizes multivariate regularly varying random vectors under polar coordinates.

Proposition 2.3.6 *The multivariate regular variation condition* (2.25) *is equivalent to the existence of a random vector* Θ *with values at the unit sphere* $\mathbb{S}^{n-1}$ *such that for all* $x > 0$

$$t I\!P\Big(\Big(\frac{||X||}{b(t)}, \frac{X}{||X||}\Big) \in \cdot\Big) \xrightarrow{v} c\nu_\alpha \cdot I\!P(\Theta \in \cdot), \quad \textit{as} \quad t \to \infty, \tag{2.27}$$

where $c > 0$, ν_α *is a measure on Borel subsets of* $(0, \infty]$ *with* $\nu_\alpha((x, \infty]) = x^{-\alpha}$, $x > 0, \alpha > 0$, *and* $||\cdot||$ *denotes an arbitrary norm in* $I\!R^n$. *We call* $S(\cdot) := I\!P(\Theta \in \cdot)$ *the spectral measure of* X.

Remark. According to Stărică (1999), p. 519, condition (2.25) of multivariate regular variation is also equivalent to

$$\frac{I\!P(||X|| > tx, X/||X|| \in \cdot)}{I\!P(||X|| > t)} \xrightarrow{v} x^{-\alpha} I\!P(\Theta \in \cdot), \quad \text{as} \quad t \to \infty, \tag{2.28}$$

where $||\cdot||$ denotes an arbitrary norm in $I\!R^n$ and, as in (2.27), we say that $S(\cdot) = I\!P(\Theta \in \cdot)$ is the *spectral measure* of X. Observe that regular variation of random variables is equivalent to regular variation of its tail functions.

Notice that Proposition 2.3.6 also implies that property (2.27) of multivariate regular variation does not depend on the choice of the norm $||\cdot||$ in $I\!R^n$. For more details regarding regular variation, O-regular variation, and multivariate regular variation we refer the reader to Bingham, Goldie, and Teugels (1987), pp. 16, pp. 61, and pp. 193 as well as to Resnick (1987), pp. 12, pp. 250.

Chapter 3

Tail dependence

The relationship between the theory of elliptically contoured distributions (copulae) and the concept of tail dependence is considered. We show that bivariate elliptical distributions (copulae) possess the so-called tail-dependence property if the tail of their generating random variable is regularly varying, and we give a necessary condition for tail dependence which is somewhat weaker than regular variation of the tail of the generating random variable. In addition, we discuss the tail-dependence property for some well-known examples of elliptical distributions, such as the multivariate normal, t, logistic, and symmetric generalized hyperbolic distributions. Further we embed the concept of tail dependence for elliptical distributions (copulae) within multivariate extreme-value theory. Finally, the latter concept is investigated for Archimedean and several other copulae.

3.1 Spherically contoured distributions

Most of the well-known market and credit risk models used in finance are based on the multivariate normal distribution; due to its tractability in portfolio calculation and simulation. However, it is well known that the normal distribution is not capable of exhibiting important properties, encountered especially in credit risk. Some of the insufficiencies of normal distributions are their light tails and the independence of extreme credit default events, where we use tail-dependence measures to identify such events. To tackle these problems we focus on the larger class of elliptically contoured distributions (in short: elliptical distributions), which includes the multivariate normal and t-distributions as representatives. Many convenient properties of the multivariate normal distribution can be extended to the class of elliptical distributions. Especially for credit risk portfolio modelling we mention the simple way of calculating the Value-at-Risk of a linear portfolio (see Section 5.2.4). We prove that the family of elliptically contoured distributions with regularly varying tails of the generating variate is almost equivalent to the class of elliptically contoured distributions possessing the dependence property for extreme credit default events (see also Schmidt (2002)).

Before we turn to elliptical distributions, we introduce the family of spherically contoured distributions (in short: spherical distributions), which forms a subclass of the family of elliptical distributions. After formulating the main results for the class of spherical distributions we can easily extend them to the more general framework of elliptical distributions. Most of the following notation is along the lines of Fang, Kotz, and Ng (1990a).

Definition 3.1.1 *Let X be an n-dimensional random vector. Then X is said to be spherically distributed if*

$$OX \stackrel{d}{=} X$$

for every orthogonal matrix $O \in \mathbb{R}^{n\times n}$.

One can easily show that X belongs to the class of spherical distributions if and only if its characteristic function $\phi_n(t),\ t \in \mathbb{R}^n$ has the form

$$\phi_n(t) = \Phi_n(t't)$$

for some function $\Phi_n : \mathbb{R} \to \mathbb{R}$. Hence, we denote the class of n-dimensional spherical distributions induced by some Φ_n by $S_n(\Phi)$ and call Φ_n the characteristic generator. If it is clear from the context we drop the subscript n.

Remark. Due to its symmetry, the characteristic function of a spherical distribution is real valued.

Besides the above characterization of spherical distributions, Schoenberg (1938) introduced the following stochastic representation, which we frequently use in the sequel. Denote by

$$\Omega_n = \{\Phi_n(\cdot) : \Phi_n(t_1^2 + \cdots + t_n^2) \text{ is an } n\text{-dimensional characteristic function}\}$$

the set of characteristic generators for n-dimensional spherical distributions.

Theorem 3.1.2 *Suppose that X is spherically distributed with characteristic generator $\Phi_n \in \Omega_n$. Then X has the representation*

$$X \stackrel{d}{=} R_n U^{(n)}, \tag{3.1}$$

where the random variable $R_n \geq 0$ is independent of the n-dimensional random vector $U^{(n)}$ which is uniformly distributed on the unit sphere $\mathbb{S}_2^{n-1}$ in $\mathbb{R}^n$.

The stochastic representation (3.1) will be a central issue of this thesis, and therefore we state the proof for completeness.

Proof. We begin by proving the following equivalence. The function Φ_n is a characteristic generator, i.e., $\Phi_n \in \Omega_n$, if and only if

$$\Phi_n(x) = \int_0^\infty \Psi_n(xr^2)\, dF_{R_n}(r), \tag{3.2}$$

where $F_{R_n} : [0,\infty) \to [0,1]$ is the distribution function of some random variable R_n and

$$\Psi_n(y'y) = \int_{x'x=1} e^{iy'x} dS_n(x)/S_n, \quad y \in I\!R^n. \tag{3.3}$$

Here $S_n(\cdot)$ denotes the Lebesgue measure on the unit sphere $\mathbb{S}_2^{n-1}$ in $I\!R^n$ and S_n denotes the surface area of the unit sphere in $I\!R^n$. Observe that $\Psi_n(y'y)$ is the characteristic function of $U^{(n)}$.

i) Let $\Phi_n \in \Omega_n$ and $x \in I\!R^n$ be arbitrary with $x'x = \|x\|^2 = 1$. Using the notation $x = (x_1, \ldots, x_n)$ and $t = (t_1, \ldots, t_n)$, we can write $g(t_1, \ldots, t_n) := \Phi_n(t't) = \Phi_n\big((\|t\|x)'(\|t\|x)\big) = g(\|t\|x_1, \ldots, \|t\|x_n)$, where $g(t_1, \ldots, t_n)$ is the characteristic function of some random vector Y with distribution function H. Then

$$\begin{aligned}
\Phi_n(t't) &= \frac{1}{S_n}\int_{\|x\|=1} g(\|t\|x_1, \ldots, \|t\|x_n)\, dS_n(x) \\
&= \frac{1}{S_n}\int_{\|x\|=1} \left[\int_{I\!R^n} e^{iy'(\|t\|x)} dH(y)\right] dS_n(x) \qquad (3.4)\\
&= \int_{I\!R^n} \left[\frac{1}{S_n}\int_{\|x\|=1} e^{i(\|t\|y)'x} dS_n(x)\right] dH(y) = \int_{I\!R^n} \Psi_n(\|t\|^2\|y\|^2)\, dH(y).
\end{aligned}$$

With $F_{R_n}(u) := \int_{\|y\|<u} dH(y)$ we obtain $\Phi_n(x) = \int_0^\infty \Psi_n(xu^2) dF_{R_n}(u)$.

ii) Let $\Phi_n(x)$ have representation (3.2) and let R_n be a random variable with distribution function F_{R_n} which is independent of $U^{(n)}$. Then

$$\begin{aligned}
I\!E(e^{it'R_nU^{(n)}}) &= \int_0^\infty I\!E(e^{irt'U^{(n)}})\, dF_{R_n}(r) \\
&= \int_0^\infty \Psi_n(\|t\|^2 r^2)\, dF_{R_n}(r) = \Phi_n(\|t\|^2) = \Phi_n(t't),
\end{aligned}$$

using the independence property of $U^{(n)}$ and R_n. Hence, $\Phi_n \in \Omega_n$.

iii) Combining part i) and ii), we obtain that

$$I\!E(e^{it'R_nU^{(n)}}) = \Phi_n(t't)$$

for some nonnegative random variable R_n independent of $U^{(n)}$ and some function Φ_n if and only if $\Phi_n \in \Omega_n$. Finally the uniqueness theorem for characteristic functions yields the theorem. □

Definition 3.1.3 *Let $X \in S_n(\Phi)$. Referring to Theorem 3.1.2 we call R_n the generating random variable or generating variate, and its distribution function F_{R_n} the generating distribution function. Further, we denote the univariate marginal distribution function of $X \in S_n(\Phi)$ by G.*

Note that each one-dimensional margin of X has the same distribution function G, due to the symmetric form of the characteristic function of X. Before we go into details let us first state the main results for spherical distributions.

Theorem 3.1.4 *Let $X \in S_n(\Phi)$, $n \geq 2$, with stochastic representation $X \stackrel{d}{=} R_n U^{(n)}$.*
i) Suppose F_{R_n}, the distribution function of R_n, has a regularly varying tail. Then all bivariate margins of X have the tail-dependence property.
ii) If X has a tail dependent bivariate margin, then the tail function of the univariate margins $\overline{G} = 1 - G$ is O-regularly varying.
iii) If X has a tail dependent bivariate margin, then the tail function $\overline{F}_{R_n} = 1 - F_{R_n}$ of R_n is O-regularly varying.

Remarks.

1. Moreover, in the framework of Theorem 3.1.4, we prove that all bivariate margins of X possess the tail-dependence property if the distribution function G of the univariate margins has a regularly varying tail, i.e., $\overline{G} = 1 - G \in RV_{-\alpha}$ with $\alpha > 0$, and $x^\alpha \overline{F}_{R_n}(x)$ satisfies the Tauberian Condition 3.1.8 (see below).
2. Later on we will state a similar theorem regarding densities, where we can prove a stronger result.

In the proof of Theorem 3.1.4 we need several lemmas and propositions, which yield interesting results themselves. First we investigate the marginal characteristics of spherical distributions.

Lemma 3.1.5 *Let $X \in S_n(\Phi)$ and $(X^{(m)}, X^{(n-m)})'$ be a partition of $X \stackrel{d}{=} R_n U^{(n)}$ with $X^{(m)} \in I\!R^m$ and $X^{(n-m)} \in I\!R^{n-m}$. Then*

$$X^{(m)} \in S_m(\Phi) \tag{3.5}$$

and

$$(X^{(m)}, X^{(n-m)})' \stackrel{d}{=} (R_n D_1 U^{(m)}, R_n D_2 U^{(n-m)})', \tag{3.6}$$

where D_1^2 is Beta$(\frac{m}{2}, \frac{n-m}{2})$ distributed and $D_2^2 = 1 - D_1^2$. Further $U^{(m)}$, $U^{(n-m)}$, and (D_1^2, D_2^2) are mutually independent random vectors.

For a *proof* of Lemma 3.1.5 we refer the reader to Fang, Kotz, and Ng (1990a).

Remark. Statement (3.5) and the symmetry of the characteristic function of spherical distributions imply that all margins of X are spherically distributed with the same characteristic generator.

The following lemma describes the relationship between the margins of a spherically distributed X and its generating variate.

Let $X \in S_n(\Phi)$ and $X^{(m)} \in S_m(\Phi)$, $1 \leq m \leq n$, be the corresponding marginal random vector. Furthermore, let $R_m \geq 0$, $1 \leq m \leq n$, denote the random variable which relates to the stochastic representation

$$X^{(m)} \stackrel{d}{=} R_m U^{(m)}.$$

Lemma 3.1.6 *If $X \in S_n(\Phi)$ and $X^{(m)} \in S_m(\Phi)$, $1 \le m \le n$, have generating variates R_n and R_m, respectively, then*

$$R_m \stackrel{d}{=} R_n B_m,$$

where the random variable B_m with $0 \le B_m \le 1$ is independent of R_n, and B_m^2 has a Beta$(\frac{m}{2}, \frac{n-m}{2})$ distribution for $1 \le m < n$ and $B_n^2 \equiv 1$.

Proof. The assertion follows immediately by Lemma 3.1.5 setting $B_m := D_1$ and realizing that $R_m = \|R_m U^{(m)}\| \stackrel{d}{=} \|R_n B_m U^{(m)}\| = R_n B_m$. □

3.1.1 Regular variation properties

An immediate consequence of Lemma 3.1.6 is that a generating variate R_n with regularly varying tail implies a regularly varying tail of R_m, which is the generating variate of the corresponding m-dimensional margin.

Proposition 3.1.7 *Let F_{R_n} and F_{R_m} be the distribution functions of the generating variate R_n and R_m corresponding to $X \in S_n(\Phi)$ and $X^{(m)} \in S_m(\Phi)$, $1 \le m \le n$. If F_{R_n} has a regularly varying tail, i.e., $\overline{F}_{R_n} \in RV_{-\alpha}$, $\alpha > 0$, then $\overline{F}_{R_m}$ is also regularly varying with the same index, i.e., $\overline{F}_{R_m} \in RV_{-\alpha}$.*

Proof. Recall that $\Phi_n \in \Omega_n$ if and only if

$$\Phi_n(x) = \int_0^\infty \Psi_n(xr^2) dF_{R_n}(r),$$

where $F_{R_n} : \mathbb{R}_+ \to [0,1]$ is a distribution function and $\Psi_n(y'y) = \int_{x'x=1} e^{iy'x} dS_n(x)/S_n$ (see Formula (3.3)). Let $\Phi_n \in \Omega_n$, then $\Phi_n \in \Omega_m$, for all $1 \le m \le n$. Consequently, for any fixed m, $1 \le m \le n$, there exists a distribution function F_{R_m} such that

$$\Phi_n(x) = \int_0^\infty \Psi_m(xr^2) dF_{R_m}(r), \quad \text{and} \quad \Psi_m(y'y) = \int_{x'x=1} e^{iy'x} dS_m(x)/S_m.$$

Suppose F_{R_n} has a regularly varying tail, i.e., $\overline{F}_{R_n} \in RV_{-\alpha}$, $\alpha > 0$. Applying Lemma 3.1.6 we obtain

$$\overline{F}_{R_m}(u) = \int_0^1 \overline{F}_{R_n}\Big(\frac{u}{y}\Big) dF_{B_m}(y), \tag{3.7}$$

where F_{B_m} denotes the distribution function of B_m and B_m^2 has a $Beta(\frac{m}{2}, \frac{n-m}{2})$ distribution for $1 \le m < n$. Moreover, utilizing the uniform convergence Theorem 2.3.2 for regularly varying functions yields that for every $\varepsilon \in (0, \alpha]$, $\varepsilon < 1$, there exists a constant $K_\varepsilon \ge 0$ such that

$$\begin{aligned} \frac{\overline{F}_{R_m}(u)}{\overline{F}_{R_n}(u)} &= \int_0^1 \frac{\overline{F}_{R_n}(\frac{u}{y})}{\overline{F}_{R_n}(u)} dF_{B_m}(y) \le \int_0^1 \Big(\frac{1}{y}\Big)^{-\alpha+\varepsilon} (1+\varepsilon) dF_{B_m}(y) \\ &= (1+\varepsilon) \int_0^1 y^{\alpha-\varepsilon} dF_{B_m}(y) = (1+\varepsilon) \mathbb{E} B_m^{\alpha-\varepsilon} < \infty \end{aligned}$$

for all $u \geq K_\varepsilon$. Similarly, there exists $\hat{K}_\varepsilon \geq 0$ such that

$$\frac{\overline{F}_{R_m}(u)}{\overline{F}_{R_n}(u)} \geq (1-\varepsilon)\mathbb{E}B_m^{\alpha+\varepsilon} > 0$$

for all $u \geq \hat{K}_\varepsilon$. Since $\varepsilon > 0$ is arbitrary and $\lim_{\varepsilon\to 0} \mathbb{E}B^{\alpha\pm\varepsilon} = \mathbb{E}B^\alpha$, this implies

$$\lim_{u\to\infty} \frac{\overline{F}_{R_m}(u)}{\overline{F}_{R_n}(u)} = \mathbb{E}B^\alpha, \quad \alpha > 0.$$

Therefore, for all $t > 0$

$$\lim_{u\to\infty} \frac{\overline{F}_{R_m}(tu)}{\overline{F}_{R_m}(u)} = \lim_{u\to\infty} \frac{\overline{F}_{R_m}(tu)}{\overline{F}_{R_n}(tu)} \frac{\overline{F}_{R_n}(u)}{\overline{F}_{R_m}(u)} \frac{\overline{F}_{R_n}(tu)}{\overline{F}_{R_n}(u)} = t^{-\alpha}$$

and we can conclude that $\overline{F}_{R_m} \in RV_{-\alpha},\ \alpha > 0,\ 1 \leq m \leq n.$

□

Remark. Similarly, one can show that the density function $f_{R_m},\ 1 \leq m \leq n$, corresponding to F_{R_m} is regularly varying with index $-\alpha - 1$, i.e., $f_{R_m} \in RV_{-\alpha-1},\ \alpha > 0$, if the density function f_{R_n} of F_{R_n} is regularly varying with the same index, i.e., $f_{R_n} \in RV_{-\alpha-1},\ \alpha > 0$.

The converse of Proposition 3.1.7 is also true under a further condition.

Condition 3.1.8 (Tauberian condition, Bingham, Goldie, and Teugels (1987), p. 197) *Let $h : \mathbb{R}_+ \to \mathbb{R}$ be a measurable function. We say that h satisfies a Tauberian condition of slow decrease if*

$$\lim_{\lambda\to 1^+} \liminf_{x\to\infty} \inf_{t\in[1,\lambda]} \{h(tx) - h(x)\} \geq 0.$$

Remark. The above condition is equivalent to h being a slowly decreasing function (see Bingham, Goldie, and Teugels (1987), p.41).

Proposition 3.1.9 *Let F_{R_n} and F_{R_m} be the distribution functions of the generating variate R_n and R_m corresponding to $X \in S_n(\Phi)$ and $X^{(m)} \in S_m(\Phi)$, respectively, $1 \leq m \leq n$. If F_{R_m} has a regularly varying tail for some $m \in \{1, \ldots, n\}$, i.e., $\overline{F}_{R_m} \in RV_{-\alpha},\ \alpha > 0$, and $x^\alpha \overline{F}_{R_n}(x)$ satisfies the Tauberian condition* 3.1.8, *then $\overline{F}_{R_n}$ is also regularly varying with the same index, i.e., $\overline{F}_{R_n} \in RV_{-\alpha}$.*

Proof. For this converse statement relative to Proposition 3.1.7 we need a deeper auxiliary result. Assume that F_{R_m} has a regularly varying tail, i.e., $\overline{F}_{R_m} \in RV_{-\alpha},\ \alpha >$

$0,\ 1 \le m < n$. Again, with Lemma 3.1.6 we obtain

$$\begin{aligned}\overline{F}_{R_m}(u) &= \int_0^1 \overline{F}_{R_n}\Big(\frac{u}{y}\Big) f_{B_m}(y)dy \\ &= \int_u^\infty \overline{F}_{R_n}(x) f_{B_m}\Big(\frac{u}{x}\Big)\frac{u}{x^2}dx \\ &= \int_0^\infty f_{B_m}\Big(\frac{u}{x}\Big)\frac{u}{x}\mathbf{1}_{(0,1)}\Big(\frac{u}{x}\Big)\overline{F}_{R_n}(x)\frac{dx}{x} = k \overset{M}{*} \overline{F}_{R_n}(u),\end{aligned}$$

where f_{B_m} is the density function of B_m and $k \overset{M}{*} \overline{F}_{R_n}$ is the Mellin convolution (see Definition 2.3.4) with kernel $k(t) = f_{B_m}(t)t\mathbf{1}_{(0,1)}(t),\ t > 0$. Since we want to transfer the regular variation property of $k \overset{M}{*} \overline{F}_{R_n}$ towards $\overline{F}_{R_n}$ we arrive at the Wiener-Tauberian theorem for Mellin convolutions (see Bingham, Goldie, and Teugels (1987), Theorem 4.8.3, p. 230). The latter theorem requires us to consider the Mellin transform of k, i.e.,

$$\check{k}(z) = \int_0^\infty t^{-z+1} f_{B_m}(t)\mathbf{1}_{(0,1)}(t)\frac{dt}{t}.$$

which converges for $-\infty < \mathrm{Re}(z) < m$, since B_m^2 is $Beta\big(\frac{m}{2}, \frac{n-m}{2}\big)$ distributed, and therefore

$$f_{B_m}(t) = \frac{2}{Beta\big(\frac{m}{2}, \frac{n-m}{2}\big)} t^{m-1}(1-t^2)^{\frac{n-m}{2}-1},\ \ t > 0,$$

and

$$\begin{aligned}\check{k}(z) &= \frac{2}{Beta\big(\frac{m}{2}, \frac{n-m}{2}\big)} \int_0^1 t^{-z+m-1}(1-t^2)^{\frac{n-m}{2}-1}dt \\ &= \frac{Beta\big(\frac{-z+m}{2}, \frac{n-m}{2}\big)}{Beta\big(\frac{m}{2}, \frac{n-m}{2}\big)}. \qquad (3.8)\end{aligned}$$

The Wiener-Tauberian theorem implies a regularly varying tail function $\overline{F}_{R_n}$, i.e., $\overline{F}_{R_n} \in RV_{-\alpha},\ \alpha > 0$, if the following two conditions hold:

i) the Wiener condition on the kernel: $\check{k} \neq 0$ for $\mathrm{Re}(z) = -\alpha$, and

ii) the Tauberian condition:

$$\lim_{\lambda \to 1^+} \liminf_{x\to\infty} \inf_{t\in[1,\lambda]} \left\{ \frac{\overline{F}_{R_n}(tx)}{(tx)^{-\alpha}} - \frac{\overline{F}_{R_n}(x)}{x^{-\alpha}} \right\} \ge 0.$$

Since we assume that $x^\alpha \overline{F}_{R_n}$ fulfills the Tauberian condition it remains to show that the Wiener condition holds.
Recall from equation (3.8) that

$$\check{k}(z) = \frac{Beta\big(\frac{-z+m}{2}, \frac{n-m}{2}\big)}{Beta\big(\frac{m}{2}, \frac{n-m}{2}\big)},\ \ -\infty < \mathrm{Re}(z) < m,\ 1 \le m < n,$$

and therefore $\check{k}(z) \neq 0$ for $-\infty < \text{Re}(z) < m,\ 1 \leq m < n$, since $\Gamma(\bar{z}) > 0$ for $\text{Re}(\bar{z}) > 0$, according to Freitag and Busam (1993), Theorem 1.10, p. 199. □

Remark. We were not able to show whether one can drop the Tauberian condition in Proposition 3.1.9.

Proposition 3.1.10 *Let* $X \stackrel{d}{=} R_n U^{(n)} \in S_n(\Phi)$. *Suppose the distribution function of* R_n *has a regularly varying tail, i.e.,* $\overline{F}_{R_n} \in RV_{-\alpha},\ \alpha > 0$. *Then the tail function* $\overline{G}$ *of the univariate margins of* X *is also regularly varying with the same index.*

Proof. Due to the symmetry of spherical distributions we may identify G as the distribution function of all univariate margins of $X \in S_n(\Phi)$. Recall that the univariate margin $X^{(1)}$ possesses the stochastic representation $X^{(1)} \stackrel{d}{=} R_1 U^{(1)}$, where $U^{(1)}$ is independent of R_1 and $U^{(1)}$ is Bernoulli distributed. Then for all $t > 0$

$$\begin{aligned} \lim_{x\to\infty} \frac{\overline{G}(tx)}{\overline{G}(x)} &= \lim_{x\to\infty} \frac{\mathbb{P}(X^{(1)} > tx)}{\mathbb{P}(X^{(1)} > x)} \\ &= \lim_{x\to\infty} \frac{\mathbb{P}(R_1 U^{(1)} > tx)}{\mathbb{P}(R_1 U^{(1)} > x)} = \lim_{x\to\infty} \frac{\mathbb{P}(R_1 > tx)}{\mathbb{P}(R_1 > x)} = t^{-\alpha}, \end{aligned}$$

because R_1 inherits the regular variation property from R_n by Proposition 3.1.7. The last but one equality follows from the independence of $U^{(1)}$ and $R_1 \geq 0$, and the Bernoulli distribution of $U^{(1)}$. □

It is obvious from the definition of spherical distributions that if $X \in S_n(\Phi)$ possesses a density, it must be of the form $g(t't) = g_n(t't)$, for some measurable function $g_n : \mathbb{R}_+ \to \mathbb{R}_+$ with $\mathbb{R}_+ = [0, \infty)$. Again, if it is clear from the context we drop the subscript n.

Definition 3.1.11 *Suppose* $X \in S_n(\Phi)$ *possesses a density function* $g(t't)$. *Then* g *is called the* density generator *of* X *and we write* $X \in S_n(g)$.

By the next result, the density generator of a spherical distribution is related to the density of its generating variate.

Lemma 3.1.12 *Let* $X \in S_n(\Phi)$ *with stochastic representation* $X \stackrel{d}{=} R_n U^{(n)}$. *Then* X *possesses a density generator* g *if and only if* R_n *has a density* f_{R_n}. *The relationship between* f_{R_n} *and* g *is given by*

$$f_{R_n}(x) = \frac{2\pi^{n/2}}{\Gamma(n/2)} x^{n-1} g(x^2), \qquad x \geq 0. \tag{3.9}$$

Proof. First, we assume that X has a density generator g. We apply the Liouville/Dirichlet integral formula (see e.g. Whittaker and Watson (1996), Ch. 12.5) which holds for any nonnegative measurable function v :

$$\int_{\mathbb{R}^n} v\Big(\sum_{i=1}^n x_i^2\Big)dx_1\dots dx_n = \frac{\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty y^{n/2-1}v(y)dy.$$

Thus, for every measurable function $h \geq 0$ we deduce

$$\begin{aligned}\mathbb{E}h(R_n) &= \mathbb{E}h(\|X\|) = \int_{\mathbb{R}^n} h(\sqrt{x'x})g(x'x)dx \\ &= \frac{\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty h(\sqrt{y})g(y)y^{n/2-1}dy = \frac{2\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty h(r)g(r^2)r^{n-1}dr.\end{aligned}$$

Setting $h(r) = \mathbf{1}_{[0,x]}(r)$ for $x \geq 0$, yields equation (3.9). For the converse we assume that R_n has density f_{R_n} and $X \in S_n(\Phi)$ with stochastic representation $X \stackrel{d}{=} R_n U^{(n)}$.Then

$$\begin{aligned}\mathbb{E}e^{it'X} &= \int_0^\infty \mathbb{E}e^{i(rt)'U^{(n)}} f_{R_n}(r)dr \\ &= \frac{\Gamma(n/2)}{\pi^{n/2}}\int_0^\infty \frac{\pi^{n/2}}{2\Gamma(n/2)u^{\frac{n-1}{2}}}u^{n/2-1}\mathbb{E}e^{i(\sqrt{u}t)'U^{(n)}} f_{R_n}(\sqrt{u})du \\ &= \frac{\Gamma(n/2)}{2\pi^{n/2}}\int_{\mathbb{R}^n} \mathbb{E}e^{i(\|x\|t)'U^{(n)}}\frac{f_{R_n}(\|x\|)}{\|x\|^{n-1}}dx_1\dots dx_n \\ &= \int_{\mathbb{R}^n} \Psi_n(\|x\|^2\|t\|^2)\frac{\Gamma(n/2)}{2\pi^{n/2}}\frac{f_{R_n}(\|x\|)}{\|x\|^{n-1}}dx_1\dots dx_n \\ &= \int_{\mathbb{R}^n} e^{it'x}\frac{\Gamma(n/2)}{2\pi^{n/2}}\frac{f_{R_n}(\|x\|)}{\|x\|^{n-1}}dx_1\dots dx_n = \Phi_n(t't),\end{aligned}$$

according to equation (3.4) and to the fact that $\int_{\mathbb{R}^n}\frac{\Gamma(n/2)}{2\pi^{n/2}}\frac{f_{R_n}(\|x\|)}{\|x\|^{n-1}}dx_1\dots dx_n = 1$. The uniqueness theorem for characteristic functions proves the assertion. □

Proposition 3.1.13 *Let $X \in S_n(\Phi)$ with stochastic representation $X \stackrel{d}{=} R_n U^{(n)}$. Then X has a regularly varying density generator $g \in RV_{-(\alpha+n)/2}$, $\alpha > 0$, if and only if R_n possesses a regularly varying density $f_{R_n} \in RV_{-\alpha-1}$, $\alpha > 0$.*

Proof. The statement of Proposition 3.1.13 immediately follows from Lemma 3.1.12 with

$$\lim_{x\to\infty}\frac{f_{R_n}(xt)}{f_{R_n}(x)} = \lim_{x\to\infty} t^{n-1}\frac{g(x^2t^2)}{g(x^2)} = t^{-\alpha-1}.$$

□

Another consequence of Lemma 3.1.6 concerning O-regularly varying tail functions is stated in the next proposition.

Proposition 3.1.14 *Let $X \in S_n(\Phi)$. Suppose that for some $m_0 \in \{1,\dots,n\}$ the generating variate R_{m_0} of the margin $X^{(m_0)}$ possesses an O-regularly varying tail, i.e., $\overline{F}_{R_{m_0}} \in OR$. Then $\overline{F}_{R_m} \in OR$ for all $1 \le m \le n$, where F_{R_m} denotes the distribution function of the generating random variable R_m.*

Proof. Suppose $X_{m_0} \in S_{m_0}(\Phi)$ and $\overline{F}_{R_{m_0}} \in OR$.

i) Consider first the case $1 \le m < m_0$. Recall from Lemma 3.1.6 that $R_m \stackrel{d}{=} R_{m_0} B_m$, where R_{m_0} and B_m are independent and B_m^2 is *Beta*$(m/2,(m_0-m)/2)$ distributed. Then for all $t > 1$

$$\begin{aligned}\liminf_{x\to\infty} \frac{\overline{F}_{R_m}(tx)}{\overline{F}_{R_m}(x)} &= \liminf_{x\to\infty} \frac{\mathbb{P}(R_m > tx)}{\mathbb{P}(R_m > x)} = \liminf_{x\to\infty} \frac{\mathbb{P}(R_m > tx)}{\mathbb{P}(R_{m_0} B_m > x)} \\ &\ge \liminf_{x\to\infty} \frac{\mathbb{P}(R_m > tx)}{\mathbb{P}(R_{m_0} > x)} \ge \liminf_{x\to\infty} \frac{\mathbb{P}(R_{m_0} > tx)}{\mathbb{P}(R_{m_0} > x)} \cdot \liminf_{x\to\infty} \frac{\mathbb{P}(R_m > tx)}{\mathbb{P}(R_{m_0} > tx)}.\end{aligned}$$

Further,

$$\begin{aligned}\liminf_{x\to\infty} \frac{\mathbb{P}(R_m > tx)}{\mathbb{P}(R_{m_0} > tx)} &= \liminf_{x\to\infty} \frac{\mathbb{P}(R_{m_0} B_m > tx)}{\mathbb{P}(R_{m_0} > tx)} \\ &= \liminf_{x\to\infty} \int_0^1 \frac{\mathbb{P}(R_{m_0} > tx/b)}{\mathbb{P}(R_{m_0} > tx)} dF_{B_m}(b) \\ &\ge \int_0^1 \liminf_{x\to\infty} \frac{\mathbb{P}(R_{m_0} > tx/b)}{\mathbb{P}(R_{m_0} > tx)} dF_{B_m}(b) > 0,\end{aligned}$$

since $\liminf_{x\to\infty} \frac{\mathbb{P}(R_{m_0} > tx/b)}{\mathbb{P}(R_{m_0} > tx)} > 0$ for all $0 < b \le 1$. The last but one inequality follows from Fatou's lemma. Combining these inequalities and applying the prerequisite $\overline{F}_{R_{m_0}} \in OR$ yields the desired result.

ii) Let now $m_0 < m \le n$. Again from Lemma 3.1.6 we know that $R_{m_0} \stackrel{d}{=} R_m B_{m_0}$, where R_m and B_{m_0} are independent and $B_{m_0}^2$ is *Beta*$(m_0/2,(m-m_0)/2)$ distributed. Then for all $t > 1$

$$\begin{aligned}0 &< \liminf_{x\to\infty} \frac{\overline{F}_{R_{m_0}}(tx)}{\overline{F}_{m_0}(x)} = \liminf_{x\to\infty} \frac{\mathbb{P}(R_m B_{m_0} > tx)}{\mathbb{P}(R_m B_{m_0} > x)} \\ &\le \liminf_{x\to\infty} \frac{\mathbb{P}(R_m > tx)}{\mathbb{P}(R_m B_{m_0} > x)} = \liminf_{x\to\infty} \Big(\int_0^1 \frac{\mathbb{P}(bR_m > x)}{\mathbb{P}(R_m > tx)} dF_{B_{m_0}}(b) \Big)^{-1} \\ &= \Big(\limsup_{x\to\infty} \int_0^1 \frac{\mathbb{P}(R_m > \frac{x}{bt})}{\mathbb{P}(R_m > x)} dF_{B_{m_0}}(b) \Big)^{-1},\end{aligned}$$

where $B_{m_0}^2$ is *Beta*$(m_0/2,(m-m_0)/2)$ distributed and independent of R_m. Therefore,

$$\limsup_{x\to\infty} \int_0^1 \frac{\mathbb{P}(R_m > \frac{x}{bt})}{\mathbb{P}(R_m > x)} dF_{B_{m_0}}(b) < \infty.$$

Assume there exist some $t > 1$ and b_0 with $1 > b_0 \geq 1/t$ such that $\limsup_{x\to\infty} \frac{\mathbb{P}(R_m > \frac{x}{b_0 t})}{\mathbb{P}(R_m > x)} = \infty$. Then

$$\limsup_{x\to\infty} \frac{\mathbb{P}(R_m > \frac{x}{bt})}{\mathbb{P}(R_m > x)} \geq \limsup_{x\to\infty} \frac{\mathbb{P}(R_m > \frac{x}{b_0 t})}{\mathbb{P}(R_m > x)} = \infty$$

for all $b \in [b_0, 1]$. Hence, for all $N \in \mathbb{N}$ there exists $x_N \geq 0$ such that

$$\frac{\mathbb{P}(R_m > \frac{x_N}{bt})}{\mathbb{P}(R_m > x_N)} \geq N \text{ for all } b \in [b_0, 1]$$

and thus

$$\int_0^1 \frac{\mathbb{P}(R_m > \frac{x_N}{bt})}{\mathbb{P}(R_m > x_N)} dF_{B_{m_0}}(b) \geq N(1 - F_{B_{m_0}}(b_0)).$$

Since N can be chosen arbitrarily large and $F_{B_{m_0}}(b_0) < 1$, we conclude

$$\limsup_{x\to\infty} \int_0^1 \frac{\mathbb{P}(R_m > \frac{x}{bt})}{\mathbb{P}(R_m > x)} dF_{B_m}(b) = \infty,$$

which leads to a contradiction. Hence,

$$\limsup_{x\to\infty} \frac{\mathbb{P}(R_m > \frac{x}{bt})}{\mathbb{P}(R_m > x)} < \infty \text{ for all } b \in [\frac{1}{t}, 1)$$

or equivalently

$$\liminf_{x\to\infty} \frac{\mathbb{P}(R_m > tbx)}{\mathbb{P}(R_m > x)} > 0 \text{ for all } b \in [\frac{1}{t}, 1) \text{ and } t > 1.$$

This completes the proof. □

3.1.2 Tail dependence

We are ready to prove Theorem 3.1.4. For convenience of the reader we state the theorem again.

Theorem 3.1.4 *Let $X \in S_n(\Phi)$, $n \geq 2$, with stochastic representation $X \stackrel{d}{=} R_n U^{(n)}$.*
i) Suppose F_{R_n}, the distribution function of R_n, has a regularly varying tail. Then all bivariate margins have the tail-dependence property.
ii) If X has a tail dependent bivariate margin, then the tail function $\overline{G}$ of the univariate margins is O-regularly varying.
iii) If X has a tail dependent bivariate margin, then the tail function $\overline{F}_{R_n}$ of R_n is O-regularly varying.

Proof. i) All bivariate margins of $X \in S_n(\Phi)$ possess the same tail-dependence coefficient due to the symmetry of the characteristic function $\Phi_n(t't)$ of X. Thus, it suffices to prove tail dependence for the bivariate margin $X^{(2)} = (X_1, X_2)$. Again, due to the symmetry of the characteristic function, both univariate margins possess the same marginal distribution function G. Let F_{R_n}, the distribution function of R_n, have a regularly varying tail $\overline{F}_{R_n} \in RV_{-\alpha}$, $\alpha > 0$, i.e.,

$$\lim_{x\to\infty} \frac{\overline{F}_{R_n}(tx)}{\overline{F}_{R_n}(x)} = t^{-\alpha} \text{ for all } t > 0. \tag{3.10}$$

Then Proposition 3.1.7 implies also a regularly varying tail for F_{R_2}, where F_{R_2} denotes the distribution function of R_2 corresponding to $X^{(2)} \stackrel{d}{=} R_2 U^{(2)}$. Note that regular variation (at ∞) implies that F_{R_2} has an infinite right endpoint. Recall from Lemma 3.1.5 that $X^{(2)}$ possesses the stochastic representation $X^{(2)} = (X_1, X_2)' = (R_2 D U_1^{(1)}, R_2\sqrt{1-D^2}U_2^{(1)})'$, where $U_1^{(1)}$, $U_2^{(1)}$, R_2, and D are mutually independent, $U_1^{(1)}$, $U_2^{(1)}$ are Bernoulli distributed, and D^2 is $Beta(\frac{1}{2}, \frac{1}{2})$ distributed. We denote the distribution function of D by F_D. Then

$$\begin{aligned}
&\lim_{v\to 1^-} \mathbb{P}(X_1 > G^{-1}(v) \mid X_2 > G^{-1}(v)) = \lim_{x\to\infty} \mathbb{P}(X_1 > x \mid X_2 > x) \\
&\quad = \lim_{x\to\infty} \mathbb{P}(R_2\sqrt{1-D^2}U_1^{(1)} > x \mid R_2 D U_2^{(1)} > x) \\
&\quad = \lim_{x\to\infty} \frac{1}{2}\mathbb{P}(R_2\sqrt{1-D^2} > x \mid R_2 D > x) \\
&\quad = \lim_{x\to\infty} \frac{1}{2} \frac{\int_0^1 \mathbb{P}\Big(R_2 > \max(\frac{1}{\sqrt{1-u^2}}, \frac{1}{u})x\Big) dF_D(u)}{\int_0^1 \mathbb{P}\Big(R_2 > \frac{x}{u}\Big) dF_D(u)} \\
&\quad = \lim_{x\to\infty} \frac{1}{2} \frac{\int_{1/\sqrt{2}}^1 \overline{F}_{R_2}\Big(\frac{x}{\sqrt{1-u^2}}\Big)/\overline{F}_{R_2}(x) dF_D(u) + \int_0^{1/\sqrt{2}} \overline{F}_{R_2}\Big(\frac{x}{u}\Big)/\overline{F}_{R_2}(x) dF_D(u)}{\int_0^1 \overline{F}_{R_2}\Big(\frac{x}{u}\Big)/\overline{F}_{R_2}(x) dF_D(u)}.
\end{aligned}$$

Together with the uniform convergence theorem (see Theorem 2.3.2) we infer that for any $\varepsilon > 0$ there exists a constant $K_\varepsilon > 0$ such that for all $x \geq K_\varepsilon$

$$\mathbb{P}(X_1 > x \mid X_2 > x) \leq \left(\frac{1+\varepsilon}{1-\varepsilon}\right) \frac{1}{2} \frac{\int_{1/\sqrt{2}}^1 (1-u^2)^{\alpha/2} dF_D(u) + \int_0^{1/\sqrt{2}} u^\alpha dF_D(u)}{\int_0^1 u^\alpha dF_D(u)}.$$

Analogously, we obtain

$$\mathbb{P}(X_1 > x \mid X_2 > x) \geq \left(\frac{1-\varepsilon}{1+\varepsilon}\right) \frac{1}{2} \frac{\int_{1/\sqrt{2}}^1 (1-u^2)^{\alpha/2} dF_D(u) + \int_0^{1/\sqrt{2}} u^\alpha dF_D(u)}{\int_0^1 u^\alpha dF_D(u)}.$$

Since $\varepsilon > 0$ can be chosen arbitrarily small we finally conclude for $\alpha > 0$

$$\begin{aligned}
&\lim_{v\to 1^-} \mathbb{P}(X_1 > G^{-1}(v) \mid X_2 > G^{-1}(v)) \\
&\quad = \frac{1}{2} \frac{\int_{1/\sqrt{2}}^1 (1-u^2)^{\alpha/2} dF_D(u) + \int_0^{1/\sqrt{2}} u^\alpha dF_D(u)}{\int_0^1 u^\alpha dF_D(u)} = \lambda_U \in (0, 1]
\end{aligned}$$

and consequently X possesses the tail-dependence property.

Using the fact that D^2 is $Beta(\frac{1}{2}, \frac{1}{2})$ distributed and therefore D has density

$$f_D(u) = \frac{2(1-u^2)^{-\frac{1}{2}}}{Beta(\frac{1}{2}, \frac{1}{2})},$$

we additionally obtain the following formula for the (upper and lower) tail-dependence coefficient $\lambda := \lambda_U = \lambda_L$. The equality of the lower and upper tail-dependence coefficient is induced by the radial symmetry of elliptically contoured distributions. Moreover,

$$\lambda = \lim_{x\to\infty} \frac{\int_0^{1/\sqrt{2}} \overline{F}_{R_2}(x/u)\frac{1}{\sqrt{1-u^2}}du}{\int_0^1 \overline{F}_{R_2}(x/u)\frac{1}{\sqrt{1-u^2}}du} = \frac{\int_0^{1/\sqrt{2}} \frac{u^\alpha}{\sqrt{1-u^2}}du}{\int_0^1 \frac{u^\alpha}{\sqrt{1-u^2}}du}. \tag{3.11}$$

Before proving part ii) we first prove part iii).

iii) Assume that $X \in S_n(\Phi)$ possesses a bivariate margin with positive (upper and lower) tail-dependence coefficient λ. Hence, all bivariate margins are tail dependent with the same tail-dependence coefficient, using the same argumentation as in part i). Without loss of generality we can consider the bivariate margin $X^{(2)} = (X_1, X_2)$. Employing the same notation as in part i) tail dependence is equivalent to the existence of the following limit

$$\frac{\int_0^{1/\sqrt{2}} \overline{F}_{R_2}(x/u)\frac{1}{\sqrt{1-u^2}}du}{\int_0^1 \overline{F}_{R_2}(x/u)\frac{1}{\sqrt{1-u^2}}du} \to \lambda \in (0,1] \quad \text{as } x \to \infty.$$

Hence, there exist constants $\varepsilon > 0$ and $K_\varepsilon \geq 0$ such that $\lambda - \varepsilon > 0$ and for all $x \geq K_\varepsilon$

$$\begin{aligned} \frac{\pi}{4}\overline{F}_{R_2}(\sqrt{2}x) &\geq \int_0^{1/\sqrt{2}} \overline{F}_{R_2}(x/u)\frac{1}{\sqrt{1-u^2}}du \\ &\geq (\lambda-\varepsilon)\int_0^1 \overline{F}_{R_2}(x/u)\frac{1}{\sqrt{1-u^2}}du \\ &\geq (\lambda-\varepsilon)\int_0^1 \overline{F}_{R_2}(x/u)du \geq (\lambda-\varepsilon)\overline{F}_{R_2}(\sqrt[3]{2}x)(1-1/\sqrt[3]{2}). \end{aligned}$$

These inequalities lead to

$$\frac{\overline{F}_{R_2}(\sqrt{2}x)}{\overline{F}_{R_2}(\sqrt[3]{2}x)} \geq \frac{4(1-1/\sqrt[3]{2})(\lambda-\varepsilon)}{\pi} =: \hat{c} > 0$$

for all $x \geq K_\varepsilon$. The latter is equivalent with characterizing $\overline{F}_{R_2}$ as O-regularly varying, since $\overline{F}_{R_2}$ is monotone decreasing (see also Bingham, Goldie, and Teugels (1987), p.65, Corollary 2.0.6). Finally, Proposition 3.1.14 implies an O-regularly varying tail function $\overline{F}_{R_n}$.

ii) Suppose again that $X \in S_n(\Phi)$ possesses a bivariate tail-dependent margin. Then part iii) and Proposition 3.1.14 yield

$$\liminf_{x\to\infty} \frac{\overline{G}(tx)}{\overline{G}(x)} = \liminf_{x\to\infty} \frac{\mathbb{P}(R_1 U^{(1)} > tx)}{\mathbb{P}(R_1 U^{(1)} > x)} = \liminf_{x\to\infty} \frac{\overline{F}_{R_1}(tx)}{\overline{F}_{R_1}(x)} > 0,$$

where $R_1 U^{(1)}$ denotes the stochastic representation of $X^{(1)}$. Therefore, $\overline{G}$ is O-regularly varying. This completes the proof. □

Corollary 3.1.15 *Let $X \in S_n(\Phi)$, $n \geq 2$, with stochastic representation $X \overset{d}{=} R_n U^{(n)}$. If G has a regularly varying tail, i.e., $\overline{G} \in RV_{-\alpha}$, $\alpha > 0$, and $x^\alpha \overline{F}_{R_n}(x)$ satisfies the Tauberian condition (see Condition* 3.1.8*), where $\overline{F}_{R_n}$ denotes the tail function of the generating variate R_n, then all bivariate margins possess the tail-dependence property.*

Proof. Let $X \in S_n(\Phi)$. Suppose the corresponding one-dimensional distribution function G has a regularly varying tail with index $-\alpha$, $\alpha > 0$. Recall that the univariate margin $X^{(1)}$ possesses the stochastic representation $X^{(1)} \overset{d}{=} R_1 U^{(1)}$, where $U^{(1)}$ is independent of R_1 and $U^{(1)}$ is Bernoulli distributed. Then for all $t > 0$

$$\begin{aligned} \lim_{x\to\infty} \frac{\overline{G}(tx)}{\overline{G}(x)} &= \lim_{x\to\infty} \frac{\mathbb{P}(X^{(1)} > tx)}{\mathbb{P}(X^{(1)} > x)} \\ &= \lim_{x\to\infty} \frac{\mathbb{P}(R_1 U^{(1)} > tx)}{\mathbb{P}(R_1 U^{(1)} > x)} = \lim_{x\to\infty} \frac{\mathbb{P}(R_1 > tx)}{\mathbb{P}(R_1 > x)} = t^{-\alpha}. \end{aligned}$$

Therefore, R_1 has also a regularly varying tail $\overline{F}_{R_1}$ with index $-\alpha$, and consequently the tail function $\overline{F}_{R_n}$ inherits the same property according to Proposition 3.1.9. Finally, Theorem 3.1.4 part i) yields the desired conclusion. □

Remark. We have not shown yet whether regular variation of the generating variate is equivalent to the tail-dependence property for spherically distributed random vectors. To answer this open question one has to consider the *ratio Mercerian theorem* as e.g. discussed in Bingham and Inoue (2000). The ratio Mercerian theorem asserts that under adequate conditions,

$$\lim_{x\to\infty} \frac{\int_0^{1/\sqrt{2}} \overline{F}_{R_2}(x/u) \frac{1}{\sqrt{1-u^2}} du}{\int_0^1 \overline{F}_{R_2}(x/u) \frac{1}{\sqrt{1-u^2}} du} = \lambda \in (0, 1]$$

implies $\overline{F}_{R_2} \in RV_{-\alpha}$, $\alpha > 0$. One of the key assumptions is that α is the only zero of

$$\int_0^1 \frac{t^\alpha}{\sqrt{1-t^2}} dt \int_0^{1/\sqrt{2}} \frac{t^z}{\sqrt{1-t^2}} dt - \int_0^{1/\sqrt{2}} \frac{t^\alpha}{\sqrt{1-t^2}} dt \int_0^1 \frac{t^z}{\sqrt{1-t^2}} dt$$

in some vertical strip $a \leq \mathrm{Re}(z) \leq b$ such that $\alpha \in (a, b)$.

In Section 3.2.3 we encounter spherical distributions which are given by their density functions. Thus, one should have a result similar to Theorem 3.1.4 regarding density functions of spherically distributed random vectors. First we prove a useful lemma.

Lemma 3.1.16 *Let F be the distribution function of an absolutely continuous non-negative random variable such that $\overline{F} \in OR$ and the corresponding density function f is eventually decreasing. Further if*

$$\limsup_{x\to\infty} \frac{\overline{F}(bx)}{\overline{F}(x)} < 1 \tag{3.12}$$

for some $b > 1$, then $f \in OR$.

Proof. First we show: There exist some constants $K, c_1, c_2 > 0$ such that

$$0 < c_1 \leq \frac{xf(x)}{\overline{F}(x)} \leq c_2 < \infty \text{ for all } x \geq K. \tag{3.13}$$

i) Let $\overline{F} \in OR$. Then there exist $K, c_2 > 0$ such that $\overline{F}(2x)/\overline{F}(x) \geq \frac{2}{c_2} > 0$ and f is decreasing for all $x \geq K/2$. Further

$$1 \geq \frac{\overline{F}(x) - \overline{F}(2x)}{\overline{F}(x)} = \frac{\int_x^{2x} f(u)du}{\overline{F}(x)} \geq \frac{xf(2x)}{\overline{F}(x)}$$

for all $x \geq K/2$ and therefore

$$\frac{2xf(2x)}{\overline{F}(2x)} \leq \frac{xf(2x)}{\overline{F}(x)} \frac{2\overline{F}(x)}{\overline{F}(2x)} \leq \frac{2\overline{F}(x)}{\overline{F}(2x)} \leq c_2$$

for all $x \geq K/2$. Thus, $\dfrac{xf(x)}{\overline{F}(x)} \leq c_2 < \infty$ for all $x \geq K$.

ii) Let $\overline{F} \in OR$, with $\limsup_{x\to\infty} \frac{\overline{F}(bx)}{\overline{F}(x)} < 1$ for some $b > 1$. Then there exist $\varepsilon, K > 0$ such that $\frac{\overline{F}(bx)}{\overline{F}(x)} \leq 1 - \varepsilon$ for all $x \geq K$. Further

$$0 < \varepsilon \leq 1 - \frac{\overline{F}(bx)}{\overline{F}(x)} = \frac{\int_x^{bx} f(u)du}{\overline{F}(x)} \leq \frac{xf(x)(b-1)}{\overline{F}(x)}$$

and therefore

$$0 < c_1 := \frac{\varepsilon}{b-1} \leq \frac{xf(x)}{\overline{F}(x)}$$

for all $x \geq K$.

iii) The conclusion $f \in OR$ now immediately follows by

$$0 < \liminf_{x\to\infty} \frac{c_1\overline{F}(tx)}{tc_2\overline{F}(x)} \leq \liminf_{x\to\infty} \frac{f(tx)}{f(x)} \leq \limsup_{x\to\infty} \frac{f(tx)}{f(x)} \leq 1.$$

□

In the context of density generators, the next condition seems to be more appropriate and easier to check (3.12).

Condition 3.1.17 *Let $h : \mathbb{R}_+ \to \mathbb{R}_+$ be a measurable function eventually decreasing such that for some $\varepsilon > 0$*

$$\limsup_{x\to\infty} \frac{h(tx)}{h(x)} \le 1 - \varepsilon \qquad \text{uniformly for all } t > 1.$$

Theorem 3.1.18 *Let $X \in S_n(g)$, $n \ge 2$, be absolutely continuous with density generator g.*
i) Suppose g is a regularly varying function, i.e., $g \in RV_{-(\alpha+n)/2}$ with $\alpha > 0$, then all bivariate margins of X possess the tail-dependence property.
ii) If X has a tail dependent bivariate margin and g satisfies Condition 3.1.17, *then g is O-regularly varying.*

Proof. i) Let g be the density generator of $X \in S_n(g)$, which is supposed to be regularly varying with index $-(\alpha+n)/2$, $\alpha > 0$. Then f_{R_n} is also regularly varying with index $-\alpha - 1$, $\alpha > 0$, according to Proposition 3.1.13. Consequently, Karamata's theorem (see Bingham, Goldie, and Teugels (1987), p. 26) implies that F_{R_n} is regularly varying with index $-\alpha$, $\alpha > 0$, i.e., $F_{R_n} \in RV_{-\alpha}$. The assertion now follows by Theorem 3.1.4.

ii) Suppose $X \in S_n(g)$ possesses a tail dependent bivariate margin. Recall that in this case all bivariate margins are tail dependent with the same tail-dependence coefficient. According to Theorem 3.1.4 the distribution function G of the univariate margins must be O-regularly varying. Further g satisfies Condition 3.1.17 and the density function $g(x^2)$ of G inherits this condition, which implies the validity of (3.12) for G. Therefore, Lemma 3.1.16 implies that g is O-regularly varying. □

Remark. The (upper and lower) tail-dependence coefficient $\lambda = \lambda_U = \lambda_L$ for spherical distributions possessing a regularly varying generating variate with index $-\alpha$, $\alpha > 0$, or regularly varying density generator with index $-\alpha/2 - 1$, $\alpha > 0$, is given by

$$\lambda = \frac{\displaystyle\int_0^{1/\sqrt{2}} \frac{u^\alpha}{\sqrt{1-u^2}} du}{\displaystyle\int_0^1 \frac{u^\alpha}{\sqrt{1-u^2}} du}; \tag{3.14}$$

see also Figure 3.1 below. Formula (3.14) has been developed in the proof of Theorem 3.1.4 (see also equation (3.11)).

3.2 Elliptically contoured distributions

3.2.1 Definition and basic properties

After our discussion of the relationship of spherical distributions and tail dependence in Section 3.1 we now turn to the more general case of random vectors with elliptically contoured distributions (in short: elliptical distributions).

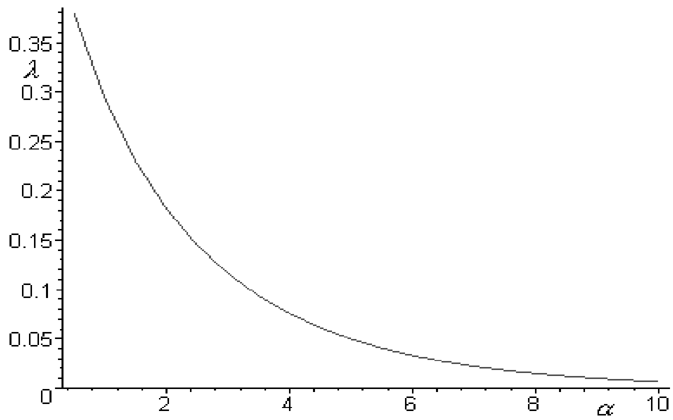


Figure 3.1: Tail-dependence coefficient λ versus regular variation index α.

Definition 3.2.1 (Elliptically contoured distribution) *Let X be an n-dimensional random vector and let $m \leq n$, $m \in \mathbb{N}$ be an arbitrary but fixed number. Then X is called elliptically distributed with parameters $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n\times n}$ if*

$$X \stackrel{d}{=} \mu + A'Y, \tag{3.15}$$

where Y is an m-dimensional spherically distributed random vector, i.e., $Y \in S_m(\Phi)$, $A \in \mathbb{R}^{m\times n}$ with $A'A = \Sigma$, and rank(Σ)=m. By $E_n(\mu, \Sigma, \Phi)$ we denote the family of all n-dimensional elliptically distributed random vectors with parameters μ, Σ, and characteristic generator Φ.

According to the stochastic representation of spherical distributions we may represent each n-dimensional elliptically distributed random vector $X \in E_n(\mu, \Sigma, \Phi)$ with parameters μ and positive-semi-definite matrix Σ, $\text{rank}(\Sigma) = m$, $m \leq n$ by

$$X \stackrel{d}{=} \mu + R_m A' U^{(m)}, \tag{3.16}$$

where $A'A = \Sigma$ and the random variable $R_m \geq 0$ is independent of the m-dimensional random vector $U^{(m)}$. The random vector $U^{(m)}$ is uniformly distributed on the unit sphere $\mathbb{S}_2^{m-1}$ in $\mathbb{R}^m$. Like in Section 3.1, the distribution function of R_m is denoted by F_{R_m}.

Remark. Sometimes the following equivalent definition of an elliptically contoured distribution is found in the literature. Let X be an n-dimensional random vector and $\Sigma \in \mathbb{R}^{n\times n}$ be a symmetric positive semi-definite matrix. If $X - \mu$, for some $\mu \in \mathbb{R}^n$, possesses a characteristic function of the form $\phi_{X-\mu}(t) = \Phi(t'\Sigma t)$ for some function $\Phi : \mathbb{R}_+ \to \mathbb{R}$, then X is said to be elliptically distributed with parameters μ (location), Σ (dispersion), and Φ. The density function, if existent, of an elliptically contoured distribution has the following form:

$$f(x) = |\Sigma|^{-1/2} g((x-\mu)'\Sigma^{-1}(x-\mu)), \quad x \in \mathbb{R}^n, \tag{3.17}$$

for some density generator function $g : \mathbb{R}_+ \to \mathbb{R}_+$. Observe that the name "elliptically contoured" distribution is related to the elliptical contours of the density f given in (3.17). For a more detailed treatment of elliptical distributions, see the monograph of Fang, Kotz, and Ng (1990a), or Cambanis, Huang, and Simons (1981).

Bingham and Kiesel (2002) (see also Section 5.2 of the present thesis) propose a semi-parametric integrative distribution model for financial modelling by estimating the parametric components (μ, Σ) and the nonparametric components (density generator g) of an elliptical distribution. Regarding the tail-dependence coefficient, we will show that it suffices to consider the tail behavior of the density generator.

Some Properties. The class of elliptically contoured distributions shares many of the structural and closure properties of multivariate normal distributions:
(i) Linear transformations of elliptically distributed random vectors are elliptically contoured. The mean vectors and scaling matrices have the following representation under linear transformations $x \mapsto Bx + b$:

$$\mu \mapsto B\mu + b, \Sigma \mapsto B\Sigma B',$$

and the density generator and characteristic generator are unchanged.
(ii) All marginal distributions of elliptical distributions are again elliptical, with the same density generator and characteristic generator (cf. Lemma 3.1.5). Thus if we partition X, μ, and Σ into

$$X = \begin{pmatrix} X^{(m)} \\ X^{(n-m)} \end{pmatrix}, \ \mu = \begin{pmatrix} \mu^{(m)} \\ \mu^{(n-m)} \end{pmatrix}, \text{ and } \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \ \Sigma_{11} \in \mathbb{R}^{m \times m}$$

and $X \in E_n(\mu, \Sigma, g)$, then also

$$X^{(m)} \in E_m(\mu^{(m)}, \Sigma_{11}, \Phi) \text{ and } X^{(n-m)} \in E_{n-m}(\mu^{(n-m)}, \Sigma_{22}, \Phi).$$

(iii) Elliptical distributions have some but not all of the useful closure properties of multivariate normal distributions under regression, or conditioning. Suppose $X, X^{(m)}, X^{(n-m)}$ are defined as above, we have partial information - we know $X^{(m)}$ - and we want to use this to infer what we can about $X^{(n-m)}$. The mean vector and covariance matrix are as in the Gaussian case:

(a) The conditional mean and covariance of $X^{(n-m)}$ given $X^{(m)}$ are

$$\bar{\mu} = \mu^{(n-m)} + \Sigma_{21}\Sigma_{11}^{-1}(X^{(m)} - \mu^{(m)}), \quad \bar{\Sigma} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12},$$

where Σ_{11}, Σ_{22}, and Σ_{21} denote to above partitions of Σ.

(b) The conditional distribution of $X^{(n-m)}$ given $X^{(m)}$ is again elliptical.

(c) The generator Φ^* of the conditional distribution is in general *not* the same as Φ:

$$X^{(n-m)} \mid (X^{(m)} = x^{(m)}) \sim E_n(\bar{\mu}, \bar{\Sigma}; \Phi^*)$$

with $\bar{\mu}$, $\bar{\Sigma}$ as above; Thus $\Phi \equiv \Phi^*$ in the Gaussian case, but not in general. See Fang, Kotz, and Ng (1990b), Ch. 2, or Embrechts, McNeil, and Straumann (2001) for further background.

3.2.2 Tail dependence

Following along the lines of Section 3.1 we first state the main result of this section before we go into details.

Theorem 3.2.2 *Let $X \in E_n(\mu, \Sigma, \Phi)$, $n \geq 2$, with positive-definite matrix Σ and stochastic representation $X \stackrel{d}{=} \mu + A'Y \stackrel{d}{=} \mu + R_n A' U^{(n)}$.*
i) If X has a tail-dependent bivariate margin, then the tail function $\overline{F}_{R_n}$ is O-regularly varying.
ii) If X has a tail-dependent bivariate margin, then the tail function $\overline{G}$ is O-regularly varying, where G denotes the distribution function of the univariate margins of Y.
iii) Suppose the distribution function F_{R_n} of R_n has a regularly varying tail. Then all bivariate margins are tail dependent.
iv) Suppose the distribution function G has a regularly varying tail, i.e., $\overline{G} \in RV_{-\alpha}$, $\alpha > 0$, and $x^\alpha \overline{F}_{R_n}(x)$ satisfies the Condition 3.1.8. Then all bivariate margins possess the tail-dependence property.

Remark. A related work by Hult and Lindskog (2002) uses different proof-techniques and obtains different results about the tail dependence property for elliptically contoured distributions. The latter authors state an equivalence of tail dependence and regular variation of the tail function $\overline{F}_{R_n}$, where we could only prove the sufficiency of regular variation. For necessity we used the theory of O-regular varying functions.

Before proving Theorem 3.2.2 we present two preliminary results. The following lemma corresponds to Lemma 3.1.6 for spherical distributions.

Lemma 3.2.3 *Suppose $X \in E_n(\mu, \Sigma, \Phi)$ with stochastic representation $X \stackrel{d}{=} \mu + R_m A' U^{(m)}$ and $X \stackrel{d}{=} \hat{\mu} + R_{\hat{m}} \hat{A}' U^{(\hat{m})}$ with $n \geq m \geq \hat{m}$. Then there exists a constant $c > 0$ such that*

$$\mu = \hat{\mu}, \quad \hat{A}'\hat{A} = cA'A, \quad R_{\hat{m}} \stackrel{d}{=} \frac{1}{\sqrt{c}} B_{\hat{m}} R_m,$$

where R_m is independent of $B_{\hat{m}}$ and $B_{\hat{m}}^2$ has a $Beta(\frac{\hat{m}}{2}, \frac{m-\hat{m}}{2})$ distribution if $m > \hat{m}$ and $B_{\hat{m}} \equiv 1$ if $m = \hat{m}$.

For a proof we refer the reader to Cambanis, Huang, and Simons (1981), pp. 372.

Lemma 3.2.4 *Let $X \in E_n(\mu, \Sigma, \Phi)$ with stochastic representation $X \stackrel{d}{=} \mu + A'Y$, $Y \in S_m(\Phi)$, $m \leq n$ $A \in \mathbb{R}^{m \times n}$, and $A'A = \Sigma$. If $\hat{A} \in \mathbb{R}^{m \times n}$ with $\hat{A}'\hat{A} = \Sigma$ then $X \stackrel{d}{=} \mu + \hat{A}'Y$.*

Proof. The lemma immediately follows by $\mathbb{E}e^{it'A'Y} = \mathbb{E}e^{i(At)'Y} = \Phi((At)'(At)) = \Phi(t'\Sigma t) = \Phi((\hat{A}t)'(\hat{A}t)) = \mathbb{E}e^{it'\hat{A}'Y}$ and the uniqueness theorem for characteristic functions. □

Proof (of Theorem 3.2.2*).* According to Lemma 3.2.3 we may restrict ourselves to 2-dimensional elliptical distributions. Moreover, without loss of generality we set $\mu = 0$, i.e., $X = (X_1, X_2)' =\in E_2(0, \Sigma, \Phi)$.

Let $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$ and $A = \begin{pmatrix} \sqrt{\sigma_{11}} & \sigma_{21}/\sqrt{\sigma_{11}} \\ 0 & \sqrt{\sigma_{22}}\sqrt{1-\rho^2} \end{pmatrix}$ the corresponding Cholesky decomposition, where $\rho := \dfrac{\sigma_{21}}{\sqrt{\sigma_{11}\sigma_{22}}} \in (-1,1)$, because Σ is positive-definite. Note that Lemma 3.2.4 justifies the consideration of this specific type of the decomposition matrix A. Observe that the generalized inverse distribution functions $F_1^{-1},\ F_2^{-1}$ of $X_1,\ X_2$ are related to each other in the following way: $\sqrt{\sigma_{22}}F_1^{-1}(u) = \sqrt{\sigma_{11}}F_2^{-1}(u)$. The cases where F_1 and F_2 have a finite right endpoint are ruled out by the regular variation or tail-dependence property. Then according to Lemma 3.1.5

$$
\begin{aligned}
&\lim_{u\to 1^-} \mathbb{P}(X_1 > F_1^{-1}(u) \mid X_2 > F_2^{-1}(u)) \\
&\quad = \lim_{x\to\infty} \mathbb{P}(X_1 > \sqrt{\sigma_{11}}x \mid X_2 > \sqrt{\sigma_{22}}x) \\
&\quad = \lim_{x\to\infty} \frac{\mathbb{P}(Y_1 > x, \frac{\sigma_{21}}{\sqrt{\sigma_{11}}}Y_1 + \sqrt{\sigma_{22}}\sqrt{1-\rho^2}Y_2 > \sqrt{\sigma_{22}}x)}{\mathbb{P}(Y_1 > x)} \\
&\quad = \lim_{x\to\infty} \frac{\mathbb{P}(R_2D > x, R_2D(\rho + \sqrt{1-\rho^2}\sqrt{\frac{1-D^2}{D^2}}U_2^{(1)}) > x)}{\mathbb{P}(R_2D > x)},
\end{aligned}
$$

where we used the mutual independence of $U_1^{(1)},\ U_2^{(1)},\ R_2$, and D, and applied the fact that $U_1^{(1)}$, and $U_2^{(1)}$ are Bernoulli distributed. Further, straightforward calculations show that

$$
\begin{aligned}
&\lim_{u\to 1^-} \mathbb{P}(X_1 > F_1^{-1}(u) \mid X_2 > F_2^{-1}(u)) \\
&\quad = \frac{1}{2}\lim_{x\to\infty}\left(\frac{\mathbb{P}(R_2D > x, R_2D(\rho + \sqrt{1-\rho^2}\sqrt{\frac{1-D^2}{D^2}}) > x)}{\mathbb{P}(R_2D > x)}\right. \\
&\qquad \left. + \frac{\mathbb{P}(R_2D > x, R_2D(\rho - \sqrt{1-\rho^2}\sqrt{\frac{1-D^2}{D^2}}) > x)}{\mathbb{P}(R_2D > x)}\right) \\
&\quad = \frac{1}{2}\lim_{x\to\infty} \frac{2\int_0^{h(\rho)} \mathbb{P}(R_2 > \frac{x}{u})\frac{1}{\sqrt{1-u^2}}du}{\int_0^1 \mathbb{P}(R_2 > \frac{x}{u})\frac{1}{\sqrt{1-u^2}}du}
\end{aligned}
$$

with $h(\rho) := \left(1 + \frac{(1-\rho)^2}{1-\rho^2}\right)^{-1/2}$. The last equality arises from the fact that D^2 is $Beta(\frac{1}{2},\frac{1}{2})$ distributed, and D and R_2 are independent random variables. Exploiting exactly the same techniques as in the spherical case for

$$
\frac{\int_0^{h(\rho)} \mathbb{P}(R_2 > \frac{x}{u})\frac{1}{\sqrt{1-u^2}}du}{\int_0^1 \mathbb{P}(R_2 > \frac{x}{u})\frac{1}{\sqrt{1-u^2}}du} \qquad \text{instead of} \qquad \frac{\int_0^{\frac{1}{\sqrt{2}}} \mathbb{P}(R_2 > \frac{x}{u})\frac{1}{\sqrt{1-u^2}}du}{\int_0^1 \mathbb{P}(R_2 > \frac{x}{u})\frac{1}{\sqrt{1-u^2}}du}
$$

yields the desired results (cf. the proof of Theorem 3.1.4 and Corollary 3.1.15). □

Remark. In the last proof we have *not* shown that an elliptically contoured random vector possesses the tail-dependence property if and only if its corresponding spherical random vector in the sense of (3.15) possesses the tail-dependence property. This still is an open question.

Theorem 3.2.5 *Let $X \in E_n(\mu, \Sigma, g)$, $n \geq 2$, with positive-definite matrix Σ and stochastic representation $X \stackrel{d}{=} \mu + A'Y \stackrel{d}{=} \mu + A'R_n U^{(n)}$, where $Y \in S_n(g)$ possesses the density generator g.*
i) If g is regularly varying, i.e., $g \in RV_{-(\alpha+n)/2}$, $\alpha > 0$, then all bivariate margins of X possess the tail-dependence property.
ii) If X possesses a tail dependent bivariate margin and g satisfies Condition 3.1.17, *then g must be O-regularly varying.*

Proof. i) Let $g \in RV_{-(\alpha+n)/2}$, $\alpha > 0$. Then, Proposition 3.1.13 and Karamata's Theorem (see Bingham, Goldie, and Teugels (1987), p. 26) imply that $\overline{F}_{R_n} \in RV_{-\alpha}$, $\alpha > 0$, where $\overline{F}_{R_n}$ denotes the tail function of the generating variate R_n of Y. Hence, all bivariate margins of X are tail dependent according to Theorem 3.2.2.

ii) Assume $X \in E_n(\mu, \Sigma, g)$ possesses a tail dependent bivariate margin. Then, according to Theorem 3.2.2, the tail function of the univariate margins of Y must be O-regularly varying. Further g satisfies Condition 3.1.17 and the density function $g(x^2)$ of G inherits this condition, which yields (3.12) for G. Therefore, Lemma 3.1.16 implies g to be O-regularly varying. □

Finally, we state a closed form expression for the (upper and lower) tail-dependence coefficient $\lambda = \lambda_U = \lambda_L$ of an elliptically contoured random vector $(X_1, X_2)' \in E_2(\mu, \Sigma, \Phi)$ with positive-definite matrix Σ having a regular varying generating variate with index $-\alpha < 0$ or having a regular varying density generator with index $-\alpha/2 - 1 < 0$:

$$\lambda = \frac{\displaystyle\int_0^{h(\rho)} \frac{u^\alpha}{\sqrt{1-u^2}} du}{\displaystyle\int_0^1 \frac{u^\alpha}{\sqrt{1-u^2}} du}, \tag{3.18}$$

with $\rho := \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$ and $h(\rho) := \left(1 + \frac{(1-\rho)^2}{1-\rho^2}\right)^{-1/2}$ (see also Figure 3.2). This formula has been derived in the proof of Theorem 3.2.2. Note that ρ corresponds to the correlation coefficient in the case of its existence (see Fang, Kotz, and Ng (1990a), p.44, for the covariance formula of elliptically contoured distributions). We remark that the (upper and lower) tail-dependence coefficient λ depends only on the (correlation) coefficient ρ and the regular variation index α.

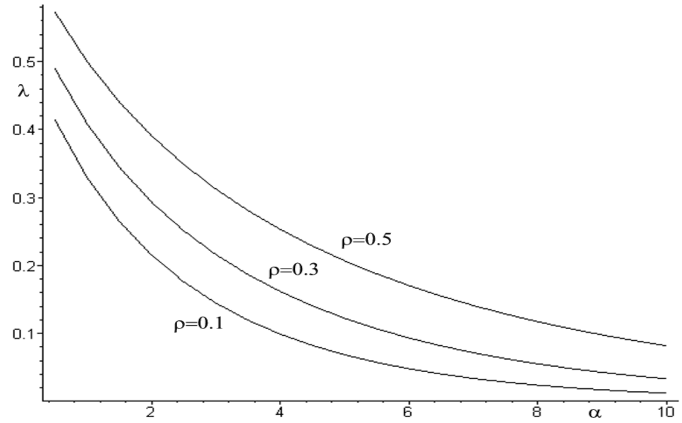


Figure 3.2: Tail-dependence coefficient λ versus regular variation index α for $\rho = 0.5,\ 0.3,\ 0.1$.

3.2.3 Examples

In this section we investigate several examples of elliptically contoured distributions. In particular we check whether they possess the tail-dependence property or their bivariate marginal distributions are tail independent, where we always assume that Σ is positive-definite.

Multivariate normal distributions and Kotz-type distributions

Multivariate normal distributions are included in the class of symmetric Kotz-type distributions. Therefore we concentrate on the latter family of distributions.

Definition 3.2.6 *Let $X \in E_n(\mu, \Sigma, g)$. Then X is called symmetrically Kotz-type distributed if the density generator g has the form*

$$g(u) = C_n u^{N-1} \exp(-ru^s), \quad r, s > 0,\ 2N + n > 2, \tag{3.19}$$

where C_n is a normalizing constant.

Theorem 3.2.7 *Let $X \in E_n(\mu, \Sigma, g)$, $n \geq 2$, be a symmetrically Kotz-type distributed random vector. Then all bivariate margins of X are tail independent.*

Proof. Observe that the density generator given in (3.19) does not belong to the class of O-regularly varying functions, because

$$\lim_{u\to\infty} \frac{g(tu)}{g(u)} = \lim_{u\to\infty} t^{N-1} \exp(-ru^s(t^s - 1)) = 0$$

for all $t > 1,\ r, s > 0$, and $2N + n > 2$. Therefore Theorem 3.2.5 yields the assertion, since its prerequisites are fulfilled. $\square$

Remark. The density generator given in (3.19) belongs to the class of multivariate normal distributions if $N = s = 1$ and $r = 1/2$.

Multivariate t-distributions and symmetric Pearson type VII distributions

Multivariate t-distributions are included in the class of symmetric Pearson-type VII distributions. Hence, we may investigate again the larger class of distributions for the tail-dependence property.

Definition 3.2.8 *Let $X \in E_n(\mu, \Sigma, g)$. Then X is called symmetrically Pearson-type VII distributed if its density generator has the form*

$$g(u) = C_n(1 + \frac{u}{m})^{-N}, \quad N > n/2,\ m > 0, \tag{3.20}$$

where C_n denotes a normalizing constant.

Theorem 3.2.9 *Let $X \in E_n(\mu, \Sigma, g)$, $n \geq 2$, be a symmetrically Pearson-type VII distributed random vector. Then all bivariate margins of X possess the tail-dependence property.*

Proof. Obviously the density generator given in (3.20) is regularly varying with index $-N$, and the assertion follows with Theorem 3.2.5. □

Remark. Setting $N = (n + m)/2$ and $m \in \mathbb{N}$ in (3.20) yields the density generator for the well-known class of multivariate t-distributions, which includes the multivariate Cauchy distribution for $m = 1$.

Multivariate logistic distributions

Definition 3.2.10 *The random vector $X \in E_n(\mu, \Sigma, g)$ is called multivariate logistically distributed if its density generator is given by*

$$g(u) = C_n \exp(-u)/(1 + \exp(-u))^2,$$

where C_n is a normalizing constant.

Theorem 3.2.11 *Suppose $X \in E_n(\mu, \Sigma, g)$, $n \geq 2$, is a logistically distributed random vector then all bivariate margins of X are tail independent.*

Proof. First, observe that $g'(u) = C_n \exp(-u)/(1 + \exp(-u))^2(\frac{2\exp(-u)}{1-\exp(-u)} - 1) < 0$ for all $u > \ln 2$. Further

$$\lim_{u\to\infty} \frac{g(tu)}{g(u)} = \lim_{u\to\infty} \exp(-u(t-1))\Big(\frac{1 + \exp(-u)}{1 + \exp(-tu)}\Big)^2 = 0$$

for all $t > 1$ and therefore Theorem 3.2.5 yields the result. □

Multivariate symmetric generalized hyperbolic distributions

Definition 3.2.12 *The random vector $X \in E_n(\mu, \Sigma, g)$ is called a multivariate symmetric generalized hyperbolic distributed random vector if its density generator is given by*

$$g(u) = C_n \frac{K_{\lambda-\frac{n}{2}}(\sqrt{\psi(\chi+u)})}{(\sqrt{\chi+u})^{\frac{n}{2}-\lambda}}, \; u > 0, \tag{3.21}$$

where $\psi, \chi > 0$, $\lambda \in I\!R$, C_n is a normalizing constant, and K_ν denotes the modified Bessel function of the third kind or MacDonald function (cf. Magnus, Oberhettinger, and Soni (1966), pp. 65).

Theorem 3.2.13 *Let $X \in E_n(\mu, \Sigma, g)$, $n \geq 2$, be a random vector with a multivariate symmetric generalized hyperbolic distribution. Then all bivariate margins of X are tail independent.*

Proof. We show that the density generator (3.21) is monotonously decreasing. Applying the following relationships for modified Bessel functions of the third kind

$$\frac{d}{dx}K_\nu(x) = -K_{\nu-1}(x) - \frac{\nu}{x}K_\nu(x), \quad K_\nu = K_{-\nu},$$

and $K_\nu(x) > 0$ for all $x \geq 0$, we obtain

$$\frac{d}{dx}K_{-\nu}(x) = \frac{d}{dx}K_\nu(x) = -K_{\nu-1}(x) - \frac{\nu}{x}K_\nu(x) < 0$$

for all $\nu \geq 0$ and $x \geq 0$. Hence, g is monotonously decreasing. Further,

$$\begin{aligned}\lim_{u\to\infty}\frac{g(tu)}{g(u)} &= \lim_{u\to\infty}\frac{K_{\lambda-\frac{n}{2}}(\sqrt{\psi(\chi+tu)})}{K_{\lambda-\frac{n}{2}}(\sqrt{\psi(\chi+u)})}\left(\sqrt{\frac{\lambda+u}{\lambda+tu}}\right)^{\frac{n}{2}-\lambda} \\ &= (\sqrt{t})^{\lambda-\frac{n}{2}}\lim_{u\to\infty}\frac{K_{\lambda-\frac{n}{2}}(\sqrt{\psi(\chi+tu)})}{K_{\lambda-\frac{n}{2}}(\sqrt{\psi(\chi+u)})} = 0\end{aligned}$$

for all $t > 1$, according to the asymptotic behavior $K_\nu \sim \sqrt{\frac{\pi}{2x}}e^{-x}(1+o(1))$ (see Abramowitz and Stegun (1964), p. 378, Formula 9.7.2), and thus

$$\lim_{x\to\infty}\frac{K_\nu(sx)}{K_\nu(x)} = \lim_{x\to\infty}\frac{1}{\sqrt{s}}e^{-x(s-1)} = 0$$

for all $s > 1$. □

3.2.4 Other dependence measures

In this section we turn to another dependence measure for elliptically contoured distributions which is useful for the study of credit and market risk models. In particular, we are interested in a dependence measure which describes, for example, the dependence structure of absolute asset returns. Then we could consider the following dependence measure for an elliptical random vector $(X_1, X_2)' \in E_2(\mu, \Sigma, \Phi)$. Let

$$\bar{\lambda} := \lim_{x \to F_2^{-1}(1)^-} \mathbb{P}(X_1 > x \mid X_2 > x)$$

if the limit exists, where F_2^{-1} denotes the generalized inverse distribution function of X_2. We call $\bar{\lambda}$ the nonstandardized tail-dependence coefficient and say $(X_1, X_2)'$ is nonstandardized tail-dependent if $\bar{\lambda} > 0$ and -independent if $\bar{\lambda} = 0$.
Analogously to Theorems 3.2.2 and 3.2.5, which deal with tail dependence, we may establish similar results for nonstandardized tail-dependence. In particular, in the case of a bivariate elliptically contoured random vector $X \in E_2(\mu, \Sigma, g)$, $n \geq 2$, Σ positive-definite, with regularly varying density generator with index $-\alpha/2 - 1$, $\alpha > 0$, we get

$$\bar{\lambda} = \frac{\frac{1}{2}\int_{h_1(\Sigma)}^{\infty} \frac{u^{-\alpha-1}}{\sqrt{u^2-1}}du + \frac{1}{2}\int_{h_2(\Sigma)}^{\infty} \frac{u^{-\alpha-1}}{\sqrt{u^2-1}}du}{\int_1^{\infty} \frac{u^{-\alpha-1}}{\sqrt{u^2-1}}du}, \tag{3.22}$$

where $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \in \mathbb{R}^{2\times 2}$,

$$h_1(\Sigma) = \sqrt{1 + \frac{(\sigma_{22} - \sigma_{12})^2}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}}, \quad \text{and} \quad h_2(\Sigma) = \sqrt{1 + \frac{(\sigma_{11} - \sigma_{12})^2}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}}.$$

Figure 3.3 shows how this dependence measure depends on the individual volatilities of X_1 and X_2.

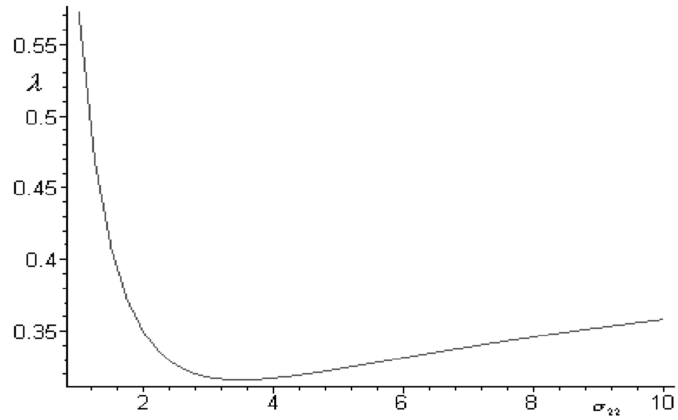


Figure 3.3: Nonstandardized tail-dependence coefficient $\bar{\lambda}$ versus volatility σ_{22}, given regular variation index $\alpha = -2.5$ and $\sigma_{11} = 2$.

Summarizing the results of this section we have found an appealing way of characterizing tail-dependent elliptically contoured distributions by regular and O-regular tail properties of the corresponding one-dimensional distribution functions or generating distributions. We applied the above results to a number of well-known elliptically contoured distributions in order to find out whether they are tail dependent or not. In this framework we showed that the symmetric Pearson-type VII distributions (including the multidimensional t-distributions) have the tail-dependence property. Therefore, there is a number of elliptically contoured distributions which inherit many useful properties of multidimensional normal distributions and moreover have additional necessary properties to model credit risk in a more realistic way. Due to the existence of good estimation and simulation techniques for Pearson-type VII or t-distributions their usage is favorable for dependence modelling within credit risk models.

3.3 Elliptically contoured copulae

3.3.1 Tail dependence

A characteristic property of elliptically contoured distributions is that all margins are elliptically distributed with the same characteristic generator or density generator, respectively. However, in many risk models, for example, we encounter the problem that the margins of multivariate asset-returns are evidently not of the same type of distribution. A possible solution regarding this problem is to join appropriate marginal distributions with a suitable copula. Here the class of elliptically contoured copulae might provide potential candidates (see also Schmidt (2003a)).

Definition 3.3.1 (Elliptically contoured copulae) *The copula C is called an elliptically contoured copula (in short: elliptical copula) if it is the copula of an elliptically contoured distribution.*

In general, if an n-dimensional copula-density c exists, then the following relationship holds between the density function f, the distribution function F, and the corresponding copula-density (cf. (2.11)):

$$f(x_1, \ldots, x_n) = c(F_1(x_1), \ldots, F_n(x_n)) \times \prod_{i=1}^{n} f_n(x_n),$$

In case the density function belongs to an elliptical distribution, it must be of the form $f(x) = |\Sigma|^{-1/2} g((x-\mu)'\Sigma^{-1}(x-\mu))$, $x \in I\!R^n$, with density generator function $g : I\!R_+ \to I\!R_+$. The name "elliptically contoured" copula is related to the elliptical contours of the latter density f.

An important issue is the estimation of copula parameters. According to Theorem 2.15 in Fang, Kotz, and Ng (1990a), elliptical copulae C which correspond to an elliptical distribution $E_n(\mu, \Sigma, \Phi)$ with positive-definite matrix $\Sigma = (\sigma_{ij})$ are uniquely determined up to a positive constant by the "correlation" matrix $\rho = (\rho_{ij})$ with $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$, $1 \le i, j \le n$, and by the characteristic generator Φ or the density generator g, respectively. Uniqueness is obtained by setting $|\Sigma| = 1$ without loss of generality. Observe that the matrix ρ contains the linear correlation coefficients only if the elliptical distribution is in the space L_2, otherwise we refer to the entries of ρ as some dependence parameters. The estimation of ρ is primarily addressed in Section 4.9.1. Nevertheless we shortly mention at this point the estimator $\hat{\rho} = (\hat{\rho}_{ij})$ proposed by Embrechts, Lindskog, and McNeil (2001) which is based on Kendall's Tau $\tau = (\tau_{ij})$ (cf. Definition 5.1.7). This estimator utilizes the relationship $\tau_{ij} = \frac{2}{\pi}\arcsin(\rho_{ij}),\ 1 \le i,j \le n$, which holds for elliptical contoured copulae. In particular, the latter authors suggest

$$\hat{\rho}_{ij} := \sin(\pi\hat{\tau}_{ij}/2) \qquad \text{with} \quad \hat{\tau}_{ij} = \frac{c_{ij} - d_{ij}}{\binom{n}{2}}, \tag{3.23}$$

where c_{ij} and d_{ij} denote the number of concordant and discordant tuples which relate to the bivariate margins of the underlying random sample. The characteristic generator Φ or the density generator g can be estimated, for example, via a nonparametric approach discussed in Bingham and Kiesel (2002) (see also Section 5.2).

Another issue arises if we restrict ourselves to a parametric elliptical copula: Which parametric elliptical copula should be chosen for which application? One criterion might be tail dependence. The discussion in Chapter 1 revealed that tail dependence represents a desired copula property especially in the context of credit-risk management. However, it was shown in the previous section that the Gaussian copula with correlation coefficient $\rho < 1$ does not inherit tail dependence which is disadvantageous. In Section 3.2.2 we showed that bivariate elliptically contoured distributions are upper and lower tail-dependent if, for instance, the tail of their density generator is regularly varying. Further, a necessary condition for tail dependence was given which is somewhat weaker than regular variation of the latter tail: precisely, the tail of the density generator must be O-regularly varying. According to Section 2.2.3 these characterizations immediately transfer to elliptical copulae because tail dependence is a copula property. Although the equivalence of tail dependence and regularly-varying density generators cannot be shown, the density generators of all well-known elliptical distributions or copulae are given either by a regularly varying tail or a not O-regularly varying tail, and therefore they yield tail dependence or tail independence. This justifies the restriction to the class of elliptical copulae with regularly varying density generator if the application requires tail dependence.

The following closed-form expressions was shown (cf. Formula 3.18) for the upper and lower tail-dependence coefficients $\lambda_{ij} := \lambda_{ij}^U = \lambda_{ij}^L,\ 1 \le i,j \le n$, of the (i,j) bivariate margin of an elliptical copula with positive-definite "correlation" matrix $\rho = (\rho_{ij})$ and regularly varying density generator g with regular-variation index $-\alpha/2-1 < 0$:

$$\lambda_{ij} = \frac{\displaystyle\int_0^{h(\rho_{ij})} \frac{u^\alpha}{\sqrt{1-u^2}}du}{\displaystyle\int_0^1 \frac{u^\alpha}{\sqrt{1-u^2}}du}, \tag{3.24}$$

where $h(\rho_{ij}) := \left(1+\frac{(1-\rho_{ij})^2}{1-\rho_{ij}^2}\right)^{-1/2}$ (see also Figure 3.2). The tail-dependence coefficient λ_{ij} depends only on the parameter ρ_{ij} and the regular-variation index α for all $1 \leq i, j \leq n$.

A large number of tail-dependence coefficients (TDCs) for various copulae are implemented in the mathematical software-program **Xplore** (see http://www.quantlet.de). The quantlet **TailCoeffCopula** derives the TDC for many bivariate copula functions. The inputs of this quantlet are the type of copula and the copula parameters. The type of copula is specified by an integer, between 1 and 34, listed in Tables 3.1, 3.2, and 3.3. For instance copula=(1) - Pareto-Clayton, copula=(4) - Gumbel, and copula=(5) - Frank in Table 3.2 and copula=(24) - t-copula in Table 3.1. The result of the quantlet is assigned to the vector `(lTDC,uTDC)` and contains the lower and upper TDC, respectively. We refer the reader to the **VARCopula** quantlet for related copula calculations.

`(lTDC,uTDC) =` **TailCoeffCopula** `(copula,parameters)`

Quantlet 2.1: Calculates the lower and upper tail-dependence coefficient for various copulae

Table 3.1 lists various elliptical distributions or copulae, the corresponding density generators (here C_n denotes a normalizing constant depending only on the dimension n) and the associated tail index α from which we easily derive the tail-dependence coefficient using formula (3.24).

3.3.2 Multivariate extreme-value theory

In this section we embed the concepts of tail dependence and elliptical copulae, introduced in Section 3.3.1, into the framework of multivariate extreme-value theory. Further we provide a different proof for the sufficient condition of an elliptical copula to be tail dependent. Recall from Section 2.1.2 that EVT is the natural choice for inferences on extreme events of random vectors or on the tail behavior of probability distributions. Usually in EVT the tail of a probability distribution is approximated by an appropriate extreme value distribution. In the one-dimensional setting the class of extreme value distributions has a solely parametric representation. Therefore it suffices to apply parametric estimation methods in contrast to nonparametric estimation methods which are less robust for small sample sizes. Multidimensional extreme value distributions are characterized by a parametric and a nonparametric component. This leads, for instance, to more complicated estimation methods. First we recall the necessary background for our purpose (see also Section 2.1.2). Let $X,\ X^{(1)},\ X^{(2)},\ldots,X^{(m)},\ m \in \mathbb{N}$ be independent multivariate random vectors with common continuous distribution function F. We say X or its distribution is in the domain of attraction of a multivariate extreme value distribution G if there exists a sequence of normalizing constants

Number & Type	Density generator g or characteristic generator Φ	Parameters	α for $n=2$
(23) Normal	$g(u) = C_n \exp(-u/2)$	-	∞
(24) t	$g(u) = C_n\Big(1+\dfrac{t}{\theta}\Big)^{-(n+\theta)/2}$	$\theta > 0$	θ
(25) Symmetric general. hyperbolic	$g(u) = C_n \dfrac{K_{\lambda-\frac{n}{2}}(\sqrt{\psi(\chi+u)})}{(\sqrt{\chi+u})^{\frac{n}{2}-\lambda}}$	$\psi, \chi > 0$ $\lambda \in \mathbb{R}$	∞
(26) Symmetric θ-stable	$\Phi(u) = \exp\Big[-\Big(\frac{1}{2}u\Big)^{\theta/2}\Big]$	$\theta \in (0,2]$	θ
(27) Logistic	$g(u) = C_n \dfrac{\exp(-u)}{(1+\exp(-u))^2}$	-	∞

Table 3.1: Tail index α for various density generators g of multivariate elliptical distributions. (K_ν denotes the modified Bessel function of the third kind (or Macdonald function) and C_n denotes a normalizing constant depending only on the dimension n).

$(a_{mi})_{m=1}^{\infty}$, $(b_{mi})_{m=1}^{\infty}$ with $a_{mi} > 0$ and $b_{mi} \in \mathbb{R}$, $i = 1, \ldots, n$, such that

$$\mathbb{P}\Big(\frac{\max_{1\le j\le m} X_1^{(j)} - b_{m1}}{a_{m1}} \le x_1, \ \ldots\ , \frac{\max_{1\le j\le m} X_n^{(j)} - b_{mn}}{a_{mn}} \le x_n\Big) \tag{3.25}$$

converges to the value $G(x_1, \ldots, x_n)$ of the limit distribution function G with non-degenerate margins as $m \to \infty$. In particular, the latter is equivalent to

$$\lim_{m\to\infty} F^m(a_{m1}x_1 + b_{m1}, \ldots, a_{mn}x_n + b_{mn}) = G(x_1, \ldots, x_n).$$

Before turning to elliptical copulae we prove the following theorem for elliptical distributions.

Theorem 3.3.2 *Let $X \in E_n(\mu, \Sigma, \Phi)$ with stochastic representation $X \stackrel{d}{=} \mu + R_n A' U^{(n)}$ and positive-definite matrix Σ. If the generating variate R_n possesses a regularly varying tail function, then X is in the domain of attraction of an extreme value distribution.*

Proof. First we show that a regularly varying R_n requires X to be in the class of multivariate regularly varying random vectors, introduced in Definition 2.3.5. Consider first the case $\mu = 0$ and $\Sigma = I$, i.e., $X \stackrel{d}{=} R_n U^{(n)}$. We need the following characterization of vague convergence stated in Resnick (1987), Proposition 3.12, p. 142: A sequence of Radon measures ν_m on some topological space $\mathbb{E}$ converges vaguely to a Radon measure ν on $\mathbb{E}$ if and only if $\lim_{m\to\infty} \nu_m(B) = \nu(B)$ for all relatively compact Borel sets $B \in \mathbb{E}$

(i.e., the closure $\overline{B}$ is compact in $\mathbb{E}$) with $\nu(\partial B) = 0$. Note that the family of Borel sets $(x, \infty] \times C,\ x > 0$, of $(0, \infty] \times \mathbb{S}_2^{n-1}$ is a generating Π-system of the class of relatively compact sets of $(0, \infty] \times \mathbb{S}_2^{n-1}$ (where $\mathbb{S}_2^{n-1}$ denotes the unit sphere in $\mathbb{R}^n$ with respect to the Euclidian norm, cf. Section 2.3.2). Thus it suffices to consider the following probabilities

$$\begin{aligned} t\mathbb{P}\Big(T\Big(\frac{X}{b(t)}\Big) \in (x, \infty] \times C\Big) &= t\mathbb{P}\Big(\Big(\frac{||X||_2}{b(t)}, \frac{X}{||X||_2}\Big) \in (x, \infty] \times C\Big) \\ &= t\mathbb{P}\Big(\frac{R_n}{b(t)} > x\Big)\mathbb{P}(U^{(n)} \in C) \to x^{-\alpha}\mathbb{P}(U^{(n)} \in C) \quad \text{as } t \to \infty, \end{aligned}$$

for $x > 0$ and $0 < b(t) \nearrow \infty$ as $t \to \infty$, since R_n and $U^{(n)}$ are stochastically independent and R_n is regularly varying. The latter convergence is locally uniform due to locally uniform convergence of regularly varying functions (see Resnick (1987), Proposition 0.5, p. 17) and the absolute continuity of $U^{(n)}$. Applying Proposition 2.3.6 yields $b(t) \nearrow \infty$ with

$$t\mathbb{P}\Big(\frac{X}{b(t)} \in \cdot\Big) \xrightarrow{v} \nu(\cdot) \tag{3.26}$$

for a Radon measure ν on $E := [-\infty, \infty]^n \backslash \{0\}$ and locally uniform convergence transfers because T and its inverse are continuous functions on E.

Let now $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$, positive-definite, be arbitrary but fixed. Consider $A \in \mathbb{R}^{n \times n}$ such that $\Sigma = A'A$ and A being a regular matrix. Define the Radon measure $\rho(\cdot) := \nu(A' \cdot)$ on E. Then

$$t\mathbb{P}\Big(\frac{X}{b(t)} \in \cdot\Big) = t\mathbb{P}\Big(\frac{A'R_nU^{(n)}}{b(t)} + \frac{\mu}{b(t)} \in \cdot\Big) \xrightarrow{v} \rho(\cdot) \tag{3.27}$$

because of the locally uniform convergence property. Further, the set

$$A'B := \{A'x \mid x \in B \text{ relatively compact in } E\}$$

is relatively compact on E and $\nu(\partial(A'B)) = 0$ if $\rho(\partial B) = \nu(A'\partial B) = 0$ since $\partial(A'B) \subset A'(\partial B)$ for regular matrices A' (Notice that $(A')^{-1}x$ is a continuous function on E). Again, Proposition 2.3.6 yields

$$t\mathbb{P}\Big(\frac{||X||_2}{b(t)} > x, X/||X||_2 \in \cdot\Big) \xrightarrow{v} cx^{-\alpha}\mathbb{P}(\Theta \in \cdot),\ c > 0, \tag{3.28}$$

for some spectral measure $S(\cdot) = \mathbb{P}(\Theta \in \cdot)$ on the unit sphere $\mathbb{S}_2^{n-1}$. We refer the reader to Hult and Lindskog (2002) for explicit calculations of the spectral measure with respect to the Euclidian and the max-norm. Finally (3.28) and Corollary 5.18 in Resnick (1987), p. 281, imply X to be in the domain of attraction of an extreme value distribution. □

Typically, elliptically contoured distributions are given by their density function or their density generator, respectively. Thus, the next corollary turns out to be helpful.

Corollary 3.3.3 *Let $X \in E_n(\mu, \Sigma, \Phi)$ be an elliptically contoured distribution with regularly varying density generator g introduced in* (3.17). *Then X is in the domain of attraction of an extreme value distribution.*

Proof. According to Proposition 3.1.13, a regularly varying density generator implies a regularly varying density function of the generating variate R_n. In particular the latter proposition yields the existence of a density function of R_n. By Karamata's Theorem (see Bingham, Goldie, and Teugels (1987), p. 26) regular variation is transferred to the tail function of R_n. The corollary now follows by Theorem 3.3.2. □

The following calculation clarifies the relationship between the spectral measure arising from multivariate regular variation of random vectors (cf. Formula (2.27)) and extreme value distributions. Let $E := [-\infty, \infty]^n \backslash \{0\}$. According to Corollary 5.18 in Resnick (1987), p. 281, every multivariate regularly varying random vector with associated spectral measure $S(\cdot)$ is in the domain of attraction of an extreme value distribution G with representation

$$G(x) = \exp\big(-\nu([-\infty, x]^c)\big), \quad x \in \mathbb{R}^n_+, \tag{3.29}$$

where $\nu(\{x \in \mathbb{R}^n \backslash \{0\} : \ ||x|| > t, x/||x|| \in \cdot\}) = t^{-\alpha} S(\cdot)$ for some arbitrary norm $|| \cdot ||$ and $[-\infty, x]^c := \{y \in E \mid y_i > x$ for some $i = 1, \ldots, n\}$. In the literature, ν is referred to as the *exponent measure* and ν_T denotes this measure under polar coordinates $T(x) = (||x||, x/||x||)$. In particular, ν_T represents a product measure and $S(\cdot) = \nu(\{x \in \mathbb{R}^n \backslash \{0\} : \ ||x|| > 1, x/||x|| \in \cdot\})$ lives on $\{y \in E : ||y|| = 1\}$. Recall that T is a bijection on E and $T^{-1}(r, a) = ra$. Set $\overline{\mathbb{S}}^{n-1} := \mathbb{S}^{n-1} \backslash (-\infty, 0]^n$ (where $\mathbb{S}^{n-1}$ denotes the unit sphere in $\mathbb{R}^n$ with respect to an arbitrary norm, cf. Section 2.3.2). Then, for $x \in \mathbb{R}^n_+$

$$\begin{aligned}
\nu([-\infty, x]^c) &= \nu_T(T([-\infty, x]^c)) \\
&= \nu_T(\{(r, a) \in \mathbb{R}_+ \times \mathbb{S}^{n-1} \mid ra_i > x_i \text{ for some } i = 1, \ldots, n\}) \\
&= \nu_T((r, a) \in \mathbb{R}_+ \times \mathbb{S}^{n-1} \mid a \in \overline{\mathbb{S}}^{n-1},\ r > \min\{\frac{x_i}{a_i},\ i \in I_a\} =: g(a)) \\
&= \int_{\overline{\mathbb{S}}^{n-1}} \int_{g(a)}^{\infty} \frac{1}{\alpha+1} \frac{1}{r^{\alpha+1}}\, dr\, S(da) = \int_{\overline{\mathbb{S}}^{n-1}} \frac{1}{g(a)^\alpha}\, S(da) \\
&= \int_{\overline{\mathbb{S}}^{n-1}} \frac{1}{\big[\min\{\frac{x_i}{a_i}, i \in I_a\}\big]^\alpha}\, S(da) \\
&= \int_{\overline{\mathbb{S}}^{n-1}} \big[\max\{\frac{a_i}{x_i}, i \in I_a\}\big]^\alpha\, S(da),
\end{aligned}$$

where $I_a = \{j \in \{1, \ldots, n\} \mid a_j > 0\}$. We summarize the above results in the following proposition.

Proposition 3.3.4 *Let X be a multivariate regularly varying random vector according to Definition 2.3.5. Then X is in the domain of attraction of a multivariate extreme value distribution*

$$G(x) = \exp\Big(- \int_{\overline{\mathbb{S}}^{n-1}} \big[\max \{ \frac{a_i}{x_i}, i \in I_a \} \big]^{\alpha} \; S(da) \Big), \quad x \in I\!R_+^n, \tag{3.30}$$

with spectral measure $S(\cdot)$ living on the unit sphere $\mathbb{S}^{n-1}$.

In general, multidimensional extreme value distributions are also characterized by an extreme value index α and a finite measure S, which is commonly referred to as the spectral or angular measure. According to the latter proposition, a multivariate regularly varying random vector is in the domain of attraction of an extreme value distribution and its spectral measure coincides with that of Definition 2.3.5.

For the family of elliptically contoured distributions the spectral measure can be given in closed form. Especially, for an elliptical random vector $X \in E_n(0, I, \Phi)$ with stochastic representation $X \stackrel{d}{=} R_n U^{(n)}$ and R_n having a regularly varying tail function with index α, we obtain

$$G(x) = \exp\Big(- \int_{\overline{\mathbb{S}}_2^{n-1}} \big[\max \{ \frac{a_i}{x_i}, i \in I_a \} \big]^{\alpha} \; da \Big), \quad x \in I\!R_+^n, \tag{3.31}$$

with $\overline{\mathbb{S}}_2^{n-1} := \mathbb{S}_2^{n-1} \backslash (-\infty, 0]^n$ and $\mathbb{S}_2^{n-1}$ denoting the unit sphere with respect to the Euclidian norm $|| \cdot ||_2$. Thus, the spectral measure $S(\cdot)$ is proportional to the uniform distribution on the unit sphere $\mathbb{S}_2^{n-1}$. Moreover, in the bivariate setup, a straightforward calculations yield

$$\begin{aligned} G(x_1, x_2) &= \exp\Big(- \frac{1}{2\pi} \Big(\frac{\sqrt{\pi}}{2} \frac{\Gamma((1+\alpha)/2)}{\Gamma(1+\alpha/2)} \big(\frac{1}{x_1^{\alpha}} + \frac{1}{x_2^{\alpha}} \big) \\ &+ \frac{1}{x_1^{\alpha}} \int_0^{\tan^{-1}(x_2/x_1)} \cos^{\alpha} \theta \; d\theta + \frac{1}{x_2^{\alpha}} \int_{\tan^{-1}(x_2/x_1)}^{\pi/2} \sin^{\alpha} \theta \; d\theta \Big) \Big). \end{aligned} \tag{3.32}$$

Having established the connection between elliptically contoured distributions and multivariate extreme-value theory, we now turn towards the relationship between the tail-dependence coefficient, elliptical copulae, and bivariate extreme-value theory. In the following we will always assume that the bivariate random vector X is in the domain of attraction of an extreme value distribution. Recall that for a bivariate random vector X with distribution function F which is in the domain of attraction of an extreme value distribution G, there must exist constants $a_{mi} > 0$ and $b_{mi} \in I\!R$, $i = 1, 2$, so that for all $(x_1, x_2)' \in I\!R_+^2$

$$\lim_{m \to \infty} F^m(a_{m1} x_1 + b_{m1}, a_{m2} x_n + b_{m2}) = G(x_1, x_2).$$

Transforming the margins of G to so-called standard Frechét margins yields

$$\lim_{m \to \infty} F_*^m(m x_1, m x_2) = G_*(x_1, x_2), \tag{3.33}$$

where F_* and G_* are the standardized distributions corresponding to F and G, respectively, with margins $G_{*i}(x_i) = \exp(-1/x_i),\ x_i > 0,\ i = 1, 2$, and

$$F_*(x_1, x_2) = F\Big(\Big(\frac{1}{1-F_1}\Big)^{-1}(x_1), \Big(\frac{1}{1-F_2}\Big)^{-1}(x_2)\Big).$$

This standardization does not cause difficulties as shown in Resnick (1987), Proposition 5.10, p. 265. Moreover, the following continuous version of (3.33) can be shown:

$$\lim_{t\to\infty} F_*^t(tx_1, tx_2) = G_*(x_1, x_2), \tag{3.34}$$

or equivalently

$$\lim_{t\to\infty} t(1 - F_*(tx_1, tx_2)) = -\log(G_*(x_1, x_2)). \tag{3.35}$$

Summarizing the above facts, we obtain

$$\begin{aligned}
&\lim_{t\to\infty} t\Big(1 - F\Big(\Big(\frac{1}{1-F_1}\Big)^{-1}(tx_1), \Big(\frac{1}{1-F_2}\Big)^{-1}(tx_2)\Big)\Big) \\
&\quad = -\log G\Big(\Big(\frac{1}{-\log G_1}\Big)^{-1}(x_1), \Big(\frac{1}{-\log G_2}\Big)^{-1}(x_2)\Big) \\
&\quad = -\log G_*(x_1, x_2).
\end{aligned} \tag{3.36}$$

Thus the tail-dependence coefficient, if existent, can be expressed as

$$\begin{aligned}
\lambda &= \lim_{v\to 1^-} \mathbb{P}(X_1 > F_1^{-1}(v) \mid X_2 > F_2^{-1}(v)) \\
&= \lim_{v\to 1^-} \mathbb{P}(X_1 > F_1^{-1}(v), X_2 > F_2^{-1}(v))/(1-v) \\
&= \lim_{t\to\infty} t\mathbb{P}\big(X_1 > F_1^{-1}(1 - \tfrac{1}{t}), X_2 > F_2^{-1}(1 - \tfrac{1}{t})\big).
\end{aligned}$$

Further, an easy calculation provides that

$$\begin{aligned}
-\log G_*(x_1, x_2) &= \lim_{t\to\infty} t\Big(1 - \mathbb{P}\Big(X_1 \le \Big(\frac{1}{1-F_1}\Big)^{-1}(tx_1), X_2 \le \Big(\frac{1}{1-F_2}\Big)^{-1}(tx_2)\Big)\Big) \\
&= \frac{1}{x_1} + \frac{1}{x_2} - \lim_{t\to\infty} \mathbb{P}\Big(X_1 > F_1^{-1}\Big(1 - \frac{1}{tx_1}\Big), X_2 > F_2^{-1}\Big(1 - \frac{1}{tx_2}\Big)\Big)
\end{aligned}$$

and hence

$$\lambda = 2 + \log G\Big(\Big(\frac{1}{-\log G_1}\Big)^{-1}(1), \Big(\frac{1}{-\log G_2}\Big)^{-1}(1)\Big). \tag{3.37}$$

The latter equation shows how to model the tail-dependence coefficient by choosing an appropriate bivariate extreme value distribution.

For an elliptical random vector $X \in E_n(0, I, \Phi)$ with stochastic representation $X \stackrel{d}{=} R_n U^{(n)}$ and R_n having a regularly varying tail function with index α, we can derive the following formula utilizing the latter results and formula (3.32)

$$\lambda = \frac{\int_{\pi/4}^{\pi/2} \cos^\alpha \theta \; d\theta}{\int_0^{\pi/2} \cos^\alpha \theta \; d\theta}. \tag{3.38}$$

Observe that for $X \in E_n(0, I, \Phi)$ this formula coincides with (3.24) after a standard substitution.

Within the framework of copulae we can rewrite (2.14) and (3.37) to obtain

$$\lambda = 2 - \lim_{t\to\infty} t\Big(1 - C\Big(1 - \frac{1}{t}, 1 - \frac{1}{t}\Big)\Big) = 2 + \log\Big(C_G\Big(\frac{1}{e}, \frac{1}{e}\Big)\Big), \tag{3.39}$$

where C and C_G denote the copula of F and G, respectively. Using the notation of co-copulae (see Nelsen (1999), p. 29), (3.39) implies

$$\lambda = 2 - \lim_{t\to\infty} t\, C_{co}\Big(\frac{1}{t}, \frac{1}{t}\Big) = 2 + \log\Big(C_G\Big(\frac{1}{e}, \frac{1}{e}\Big)\Big). \tag{3.40}$$

The above results lead to the observation that a bivariate random vector inherits the tail-dependence property if the standardized distribution F_* or the related copula function C (which equals the copula of F) is in the domain of attraction of an extreme value distribution with dependent margins. Consequently, it is not necessary that the bivariate distribution function F itself is in the domain of attraction of some extreme value distribution. This is an important property for asset portfolio modelling.

Remark. In particular every bivariate regularly varying random vector whose spectral measure is not concentrated on $(c,0)'$ and $(0,c)'$ for some $c > 0$ possesses the tail-dependence property according to Corollary 5.25 in Resnick (1987), p. 292. This is in accordance with Theorems 3.2.5 and 3.2.2 since the spectral measure of a non-degenerated elliptical distribution is not concentrated on single points.

We finish this section with an important theorem about elliptical copulae which is based on the previous results.

Theorem 3.3.5 *Let C be an elliptical copula corresponding to an elliptically distributed random vector $X \stackrel{d}{=} \mu + R_n A' U^{(n)} \in E_n(\mu, \Sigma, \Phi)$, Σ positive-definite, with regularly varying generating variate R_n or regularly varying density generator. Then C is in the domain of attraction of some extreme value distribution and all bivariate margins of C possess the tail-dependence property.*

Proof. Theorem 3.3.2 and Corollary 3.3.3 imply X to be in the domain of attraction of some extreme value distribution. According to Proposition 5.10 in Resnick (1987), p. 265, all margins X_i, $i = 1, \ldots, n$, and the standardized distribution F_*, i.e., $F_*(x_1, \ldots, x_n) = F((1/(1-F_1))^{-1}(x_1), \ldots, (1/(1-F_n))^{-1}(x_n)$ are in the domain of attraction of some extreme value distributions and $\lim_{m\to\infty} F_*^m(mx_1, \ldots, mx_n) = G(x_1, \ldots, x_n)$. Since $F_*(x_1, \ldots, x_n) = C(1 - 1/x_1, \ldots, 1 - 1/x_n) = C_*(x_1, \ldots, x_n)$, $x_1, \ldots, x_n \geq 1$, and uniform distributions on $[0, 1]$ are in the domain of attraction of some extreme value distribution, we conclude again with Proposition 5.10 in Resnick (1987) that C is in the domain of attraction of some extreme value distribution. Applying Proposition 3.1.7, every bivariate margin of C is an elliptical copula with regularly varying generating variate or regularly varying density generator. Thus tail dependence for all bivariate margins of C follows by the results stated ahead the present theorem. □

Summarizing the results of this section, we have found that elliptical copulae provide appealing dependence structures for asset portfolio modelling within internal credit-risk management. We characterized those elliptical copulae which incorporate dependencies of extreme credit default events by the so-called tail-dependence property. Further we showed that most elliptical copulae having the tail-dependence property are in the domain of attraction of an extreme value distribution. Thus powerful tools of extreme-value theory can be applied. Moreover, the application of elliptical copulae is recommended due to the existence of good estimation and simulation techniques (see also Section 5.2).

3.4 Archimedean copulae

Archimedean copulae represent an important class of copulae which are easy to construct. Therefore we explore the tail-dependence coefficient for this family of copulae (see also Schmidt (2003b)). Referring to Härdle, Kleinow, and Stahl (2002), pp. 35, and the Xplore-`VaR` quantlib, Archimedean copulae are implemented and investigated within the Value-at-Risk framework of the mathematical software-package **Xplore**. A bivariate *Archimedean copula* has the form $C(u,v) = \phi^{[-1]}(\phi(u) + \phi(v))$ for some continuous, strictly decreasing, and convex function $\phi : [0,1] \to [0,\infty]$ such that $\phi(1) = 0$, where the pseudo-inverse function $\phi^{[-1]}$ is defined by

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t), & 0 \le t \le \phi(0), \\ 0, & \phi(0) < t \le \infty. \end{cases}$$

We call ϕ the *generator* of an Archimedean copula which is *strict* if $\phi(0) = \infty$. In that case $\phi^{[-1]} \equiv \phi^{-1}$. Within the framework of tail dependence for Archimedean copulae we provide the following theorem. Note that the one-sided derivatives of ϕ exist as ϕ is a convex function. In particular, $\phi'(1)$ and $\phi'(0)$ denote the one-sided derivatives at the domains boundary of ϕ.

Theorem 3.4.1 *Then,*

i) upper tail-dependence implies $\phi'(1) = 0$ *and* $\lambda_U = 2 - (\phi^{-1} \circ 2\phi)'(1)$,

ii) $\phi'(1) < 0$ *implies upper tail-independence,*

iii) $\phi'(0) > -\infty$ *or a non-strict* ϕ *implies lower tail-independence,*

iv) lower tail-dependence implies $\phi'(0) = -\infty$, *a strict* ϕ, *and* $\lambda_L = (\phi^{-1} \circ 2\phi)'(0)$.

Proof. ii) Let $\phi'(1) < 0$. According to Theorem 47.3 in Heuser (2000), $\left(\phi^{-1}\right)'(0)$ exists. Note that $\phi(1) = 0$. Thus

$$\begin{aligned} \lim_{v \to 1^-} \frac{1 - 2v + C(v,v)}{1 - v} &= 2 - \lim_{v \to 1^-} \frac{\phi^{-1}(2\phi(v)) - 1}{v - 1} \\ &= 2 - (\phi^{-1} \circ 2\phi)'(1) = 2 - (\phi^{-1})'(2\phi(1)) \cdot \phi'(1) \\ &= 2 - \frac{1}{\phi'(\phi^{-1}(0))} \cdot 2\phi'(1) = 0 = \lambda_U. \end{aligned}$$

i) From part ii) we observe that if C is upper tail-dependent then $\phi'(1) = 0$ and

$$\lambda_U = \lim_{v \to 1^-} \frac{1 - 2v + C(v,v)}{1 - v} = 2 - (\phi^{-1} \circ 2\phi)'(1).$$

iii) First notice that if ϕ is not strict then

$$\lambda_L = \lim_{v \to 0^+} \frac{C(v,v)}{v} = \lim_{v \to 0^+} \frac{\phi^{[-1]}(2\phi(v))}{v} = 0$$

since $\phi^{[-1]}(t) = 0$ for $t \in [\phi(0), \infty)$ and ϕ is continuous on $[0,1]$. If $\phi'(0)$ exists, then ϕ is not strict and therefore C is lower tail-independent.

iv) If ϕ is strict, i.e., $\phi(0) = \infty$, then $\phi^{[-1]} \equiv \phi^{-1}$ and

$$\lambda_L = \lim_{v \to 0^+} \frac{\phi^{-1}(2\phi(v))}{v} = (\phi^{-1} \circ 2\phi)'(0).$$

Further, note that if ϕ is strict, then $\lim_{v \to 0^+} \phi(v) = \infty$. Thus, $\phi'(0)$ does not exist. □

Table 3.2 lists various Archimedean copulae in the same ordering as in Table 2.1 in Härdle, Kleinow, and Stahl (2002), p. 42, or in Nelsen (1999), p. 94, and the corresponding upper and lower tail-dependence coefficient (TDC).

3.5 Other copulae

For many other closed-form copulae we can explicitly derive the tail-dependence coefficient. Table 3.3 lists various well-known copula functions and the corresponding lower and upper TDC. Most copulae are listed in the monograph of Joe (1997), pp. 39.

Number & Type	$C(u,v)$	$\phi_\theta(t)$	$\theta \in$	upper TDC	lower TDC
(1) Pareto	$\max\left([u^{-\theta}+v^{-\theta}-1]^{-1/\theta},0\right)$	$t^{-\theta}-1$	$[-1,\infty)\backslash\{0\}$	0 for $\theta>0$	$2^{-1/\theta}$ for $\theta>0$
(2)	$\max\left(1-\left[(1-u)^\theta+(1-v)^\theta\right]^{1/\theta},0\right)$	$(1-t)^\theta$	$[1,\infty)$	$2-2^{1/\theta}$	0
(3) Ali-Mikhail-Haq	$\dfrac{uv}{1-\theta(1-u)(1-v)}$	$\log\dfrac{1-\theta(1-t)}{t}$	$[-1,1)$	0	0
(4) Gumbel	$\exp\left(-\left[(-\log u)^\theta+(-\log v)^\theta\right]^{1/\theta}\right)$	$(-\log t)^\theta$	$[1,\infty)$	$2-2^{1/\theta}$	0
(12)	$\left(1+\left[(u^{-1}-1)^\theta+(v^{-1}-1)^\theta\right]^{1/\theta}\right)^{-1}$	$\left(\frac{1}{t}-1\right)^\theta$	$[1,\infty)$	$2-2^{1/\theta}$	$2^{-1/\theta}$
(14)	$\left(1+\left[(u^{-1/\theta}-1)^\theta+(v^{-1/\theta}-1)^\theta\right]^{1/\theta}\right)^{-\theta}$	$\left(t^{-1/\theta}-1\right)^\theta$	$[1,\infty)$	$2-2^{1/\theta}$	$\frac{1}{2}$
(19)	$\theta/\log\left(e^{\theta/u}+e^{\theta/v}-e^\theta\right)$	$e^{\theta/t}-e^\theta$	$(0,\infty)$	0	1

Table 3.2: Tail-dependence coefficient (TDC) for various Archimedean copulae. Numbers correspond to Table 4.1 in Nelsen (1999), p. 94.

Number & Type	$C(u,v)$	Parameters	upper TDC	lower TDC
(28) Raftery	$g(\min(u,v),\max(u,v);\theta)$ $g(x,y;\theta)=x-\frac{1-\theta}{1+\theta}x^{1/(1-\theta)}\Big(y^{-\theta/(1-\theta)}-y^{1/(1-\theta)}\Big)$	$\theta\in[0,1]$	0	$\dfrac{2\theta}{1+\theta}$
(29) BB1	$\Big[1+\Big\{(u^{-\theta}-1)^{\delta}+(v^{-\theta}-1)^{\delta}\Big\}^{1/\delta}\Big]^{-1/\theta}$	$\theta\in(0,\infty)$ $\delta\in[1,\infty)$	$2-2^{1/\delta}$	$2^{-1/(\theta\delta)}$
(30) BB4	$\Big[u^{-\theta}+v^{-\theta}-1-\Big\{(u^{-\theta}-1)^{-\delta}+(v^{-\theta}-1)^{-\delta}\Big\}^{-1/\delta}\Big]^{-1/\theta}$	$\theta\in[0,\infty)$ $\delta\in(0,\infty)$	$2^{-1/\delta}$	$(2-$ $2^{-1/\delta})^{-1/\theta}$
(31) BB7	$1-\Big(1-\Big[\{1-(1-u)^{\theta}\}^{-\delta}+\{1-(1-v)^{\theta}\}^{-\delta}-1\Big]^{-1/\delta}\Big)^{1/\theta}$	$\theta\in[1,\infty)$ $\delta\in(0,\infty)$	$2-2^{1/\theta}$	$2^{-1/\delta}$
(32) BB8	$\frac{1}{\delta}\Big(1-\Big[1-\{1-(1-\delta)^{\theta}\}^{-1}\cdot$ $\{1-(1-\delta u)^{\theta}\}\{1-(1-\delta v)^{\theta}\}\Big]^{1/\theta}\Big)$	$\theta\in[1,\infty)$ $\delta\in[0,1]$	$2-$ $2(1-\delta)^{\theta-1}$	0
(33) BB11	$\theta\min(u,v)+(1-\theta)uv$	$\theta\in[0,1]$	θ	θ
(34) C_Ω in Junker et al. (2002)	$\beta C^{s}_{(\bar\theta,\bar\delta)}(u,v)-(1-\beta)C_{(\theta,\delta)}(u,v)$ with Archimedean generator $\phi_{(\theta,\delta)}(t)=\Big(-\log\frac{e^{-\theta t}-1}{e^{-\theta}-1}\Big)^{\delta}$ $C^{s}_{(\bar\theta,\bar\delta)}$ is the corresp. survival copula with param. $(\bar\theta,\bar\delta)$	$\theta,\bar\theta\in I\!R\backslash\{0\}$ $\delta,\bar\delta\geq 1$ $\beta\in[0,1]$	$(1-\beta)\cdot$ $(2-2^{1/\delta})$	$\beta(2-2^{1/\bar\delta})$

Table 3.3: Tail-dependence coefficient (TDC) for various copulae (copulae BBx are given in Joe (1997)).

Chapter 4

Estimation of extremal dependence

Copulae handle the overall dependence structure of multivariate distributions whereas tail copulae particularly describe the dependence structure within the tail region of multivariate distributions. The first part of the present chapter introduces several nonparametric estimators for the tail copula. For these estimators, results on weak convergence, asymptotic normality, and strong consistency are provided by means of a functional Delta method. Furthermore, a general rank order statistics for extreme events is investigated and the relationship to the concept of tail dependence is given. The second part of the present chapter concentrates on particular problems of estimation for elliptically contoured distributions. Further, we elaborate on a guideline how to estimate the upper and lower tail-dependence coefficients for arbitrary distributions. We distinguish between parametric and nonparametric estimators as well as estimation methods which are based on the entire data sample or only on extreme data. Finally, a simulation study compares the introduced estimators and illustrates some of their statistical properties.

4.1 The tail copula

According to Section 2.2.4, tail copulae are functions that describe the dependence structure of multivariate distributions in the corresponding tail region. In other words, the concept of tail copulae deals with dependencies of extreme events of a given random vector. An important application of tail copulae in finance and actuarial sciences is the modelling of dependencies between large default events in credit portfolios or between extreme insurance claims. Similar to copula functions we may construct multivariate extreme value distributions with a given tail copula.

Concerning the estimation of copula functions, several parametric, semi-parametric, and nonparametric procedures have been proposed in the literature (cf. Genest and Rivest (1993) and Genest, Ghoudi, and Rivest (1995)). Regarding the nonparametric procedures, Deheuvels (1979, 1981) and Fermanian, Radulović, and Wegkamp (2002) establish weak convergence of the so-called empirical copula process under independent and dependent marginal distributions. In contrast to that, the first part of the present chapter provides several results concerning weak convergence of the so-called empirical tail-copula process. The latter process is utilized as a nonparametric estimator for the tail copula and the tail-dependence coefficient. For further results and details we refer to Schmidt and Stadtmüller (2002).

For completeness, we restate the definition of the tail copula and its relationship to the tail-dependence coefficient, introduced already earlier in Section 2.2.4. Throughout this chapter we denote $\bar{\mathbb{R}}^n_+ := [0,\infty]^n \backslash \{(\infty,\dots,\infty)\}$.

Definition 2.2.7 (Tail copulae) *The function* $\Lambda_U^{I,J} : \bar{\mathbb{R}}^n_+ \to \mathbb{R},\ I,J \subset \{1,\dots,n\}$, $I \cap J = \emptyset$ *is called an upper tail copula associated with the n-dimensional distribution function F if the following limit exists everywhere*

$$\Lambda_U^{I,J}(x) := \lim_{t\to\infty} \mathbb{P}(X_i > F_i^{-1}(1 - x_i/t),\ \forall i \in I \mid X_j > F_j^{-1}(1 - x_j/t),\ \forall j \in J). \quad (4.1)$$

The corresponding lower tail copula is defined by

$$\Lambda_L^{I,J}(x) := \lim_{t\to\infty} \mathbb{P}(X_i \le F_i^{-1}(x_i/t),\ \forall i \in I \mid X_j \le F_j^{-1}(x_j/t),\ \forall j \in J) \quad (4.2)$$

provided the limit exists.

For simplicity and notational convenience all further definitions and results are provided for the bivariate case first. The multidimensional extensions are given in Section 4.8. Within the bivariate framework we consider a random vector $(X,Y)'$ with distribution function F and continuous marginal distribution functions G and H. The estimation becomes easier if the following slight modification of the tail copula is utilized:

$$\Lambda_U(x,y) := y \cdot \Lambda_U^{\{1\},\{2\}}(x,y) \quad \text{and} \quad (4.3)$$

$$\Lambda_L(x,y) := y \cdot \Lambda_L^{\{1\},\{2\}}(x,y), \quad x \in \bar{\mathbb{R}}_+, y \in \mathbb{R}_+, \quad (4.4)$$

where the indices $\{1\}$ and $\{2\}$ can be dropped. Further, set $\Lambda_U(x,\infty) := x$ and $\Lambda_L(x,\infty) := x$ for all $x \in \mathbb{R}_+$.

According to Definition 2.2.4 in Section 2.2.4, upper tail-dependence (tail-independence) is equivalent with the existence of $\Lambda_U(1,1) > 0$ ($\Lambda_U(1,1) = 0$), i.e., $\lambda_U = \Lambda_U(1,1)$. Similarly, $\lambda_L = \Lambda_L(1,1)$ if existent.

Estimating the tail copula can be coped with techniques from extreme-value theory. It can be shown (see Resnick (1987), Chapter 5, and Section 2.2.4 of the present thesis) that the upper tail-copula exists and $\Lambda_U \neq 0$ if the associated distribution function F lies in the domain of attraction of an extreme value distribution with dependent margins. Similar results hold for the lower tail copula. However, the latter is only a sufficient condition. In bivariate extreme-value theory the major interest concerns the probability

$$\mathbb{P}(X > G^{-1}(1-x) \text{ or } Y > H^{-1}(1-y)), \quad (4.5)$$

where the following important approximation can be shown:

$$\mathbb{P}(X > G^{-1}(1-x/t) \text{ or } Y > H^{-1}(1-y/t)) \quad (4.6)$$
$$\approx \mathbb{P}(X > G^{-1}(1-x) \text{ or } Y > H^{-1}(1-y))/t$$

for small x and y. Hence, the set $[G^{-1}(1-x/t), 1] \times [H^{-1}(1-y/t), 1]$ is shifted to $[G^{-1}(1-x), 1] \times [H^{-1}(1-y), 1]$ for appropriate x and y so that enough observations are available and the empirical distribution function serves as a suitable estimator. In the context of (upper) tail copulae the probability under consideration closely relates to

$$\mathbb{P}(X > G^{-1}(1-x) \text{ and } Y > H^{-1}(1-y)). \tag{4.7}$$

In Theorem 4.3.1 we prove that the probability (4.7) possesses the same homogeneity property as in (4.6) if the corresponding upper tail-copula exists and the upper tail-dependence coefficient is greater than zero, i.e., $(X,Y)'$ is upper tail-dependent. In particular, $t \mapsto \mathbb{P}(X > G^{-1}(1-x/t)$ and $Y > H^{-1}(1-y/t))$ is regularly varying at infinity with index -1.

In the following we propose two types of estimators based on the empirical probabilities of (4.5) and (4.7), which utilize the homogeneity property mentioned above. Notice, in case $(X,Y)'$ is tail independent, the latter property does not hold for (4.7). Here, an adjusted homogeneity property can be obtained by assuming that $t \mapsto \mathbb{P}(X > G^{-1}(1-x/t)$ and $Y > H^{-1}(1-y/t))$ is regularly varying at infinity with index $-1/\eta$, $\eta < 1$. The parameter η was introduced by Ledford and Tawn (1997, 1996, 1998) as the coefficient of asymptotic dependence given tail independence. Several estimators for η and related tests for tail independence were introduced by Coles, Heffernan, and Tawn (1999), Peng (1999), and Draisma, Drees, Ferreira, and de Haan (2001). However, according to the latter paper the tests on tail dependence or tail independence show a disappointing behavior. In contrast to these approaches we concentrate on tail dependence (e.g. the case $\eta = 1$).

4.2 Nonparametric estimators

Suppose $(X,Y)'$, $(X^{(1)}, Y^{(1)})', \ldots, (X^{(m)}, Y^{(m)})'$ are iid bivariate random vectors with distribution function F having marginal distribution functions G, H and copula C.

In the forthcoming, two nonparametric estimators for the modified lower and upper tail copula $\Lambda_U(x,y)$ and $\Lambda_L(x,y)$, $(x,y)' \in \bar{R}^2_+$, are proposed. Let C_m denote the empirical copula defined by

$$C_m(u,v) = F_m(G_m^{-1}(u), H_m^{-1}(v)),\ (u,v)' \in [0,1]^2 \tag{4.8}$$

with F_m, G_m, H_m being the empirical distribution functions corresponding to F, G, H. Then, $C_m((a,b] \times (c,d])$ is called the empirical copula-"measure" of the interval $(a,b] \times (c,d] \subset [-\infty, \infty]^2$ given by

$$C_m((a,b] \times (c,d]) := C_m\big(((a,b] \times (c,d]) \cap [0,1]^2\big).$$

It is important to note that the choice of the empirical distribution function to model the marginal distributions avoids any misidentifications due to a wrong parametric fit of the marginal distributions. Empirical investigations regarding such misidentifications are provided in Section 4.10.

Let $R_{m1}^{(j)}$ and $R_{m2}^{(j)}$ denote the rank of $X^{(j)}$ and $Y^{(j)}$, $j = 1, \ldots, m$, respectively. The first set of estimators are based on formulae (4.1) and (4.2):

$$\begin{aligned} \hat{\Lambda}_{U,m}(x, y) &:= \frac{m}{k} C_m\Big(\Big(1 - \frac{kx}{m}, 1\Big] \times \Big(1 - \frac{ky}{m}, 1\Big]\Big) \\ &\approx \frac{1}{k} \sum_{j=1}^{m} \mathbf{1}_{\{R_{m1}^{(j)} > m - kx \text{ and } R_{m2}^{(j)} > m - ky\}} \end{aligned} \tag{4.9}$$

and

$$\hat{\Lambda}_{L,m}(x, y) := \frac{m}{k} C_m\Big(\frac{kx}{m}, \frac{ky}{m}\Big) \approx \frac{1}{k} \sum_{j=1}^{m} \mathbf{1}_{\{R_{m1}^{(j)} \le kx \text{ and } R_{m2}^{(j)} \le ky\}} \tag{4.10}$$

with some parameter $k \in \{1, \ldots, m\}$ to be chosen by the statistician. For the asymptotic results we assume throughout this chapter that $k = k(m) \to \infty$ and $k/m \to 0$ as $m \to \infty$. The estimators $\hat{\Lambda}_{U,m}(x, y)$ and $\hat{\Lambda}_{L,m}(x, y)$ are referred to as *empirical tail-copulae.*

The far right sides in equations (4.9) and (4.10) provide two approximative rank order statistics which are based on a slightly modified representation of the empirical tail-copula. This representation was proposed by Genest, Ghoudi, and Rivest (1995), i.e.,

$$\bar{C}_m(u, v) = \frac{1}{m} \sum_{j=1}^{m} \mathbf{1}_{\{G_m(X^{(j)}) \le u \text{ and } H_m(Y^{(j)}) \le v\}}, \ (u, v)' \in [0, 1]^2. \tag{4.11}$$

In the present thesis, the results about weak convergence and consistency are established for the tail-copula estimators $\hat{\Lambda}_{U,m}$ and $\hat{\Lambda}_{L,m}$. However, the following reasoning shows that all results hold also for the corresponding rank order statistics.

Note that the empirical tail-copulae and the rank order statistics coincide on the grid $\{(i/k, j/k),\ 1 \le i, j \le m\}$. Otherwise the pointwise differences are at most $2/k$. Consider e.g. the lower empirical tail-copula $\hat{\Lambda}_{L,m}(x, y)$ which is left-continuous whereas the corresponding rank order statistics is right-continuous. The difference between the latter estimators is bounded by

$$\begin{aligned} &\sup_{(x,y)' \in \bar{\mathbb{R}}_+^2} \Big|\hat{\Lambda}_{L,m}(x, y) - \frac{1}{k} \sum_{j=1}^{m} \mathbf{1}_{\{R_{m1}^{(j)} \le kx \text{ and } R_{m2}^{(j)} \le ky\}}\Big| \\ &\quad \le \max_{1 \le i,j \le m} \Big|\frac{m}{k} C_m\Big(\frac{i}{m}, \frac{j}{m}\Big) - \frac{m}{k} C_m\Big(\frac{i-1}{m}, \frac{j-1}{m}\Big)\Big| \le \frac{m}{k}\frac{2}{m} = \frac{2}{k}. \end{aligned}$$

The second pair of estimators were essentially introduced and investigated by Huang (1992), Chapter 2, Peng (1998), pp. 96, and Durrleman, Nikeghbali, and Roncalli (2000) in the context of stable tail-dependence functions. According to Section 2.2.4 of the present thesis, the relationship between the bivariate upper tail-copula and the stable tail-dependence function l is given via $\Lambda_U(x, y) = x + y - l(x, y)$. The latter authors discuss the function l with respect to questions arising from EVT. They show consistency and asymptotic normality for the estimator below. However, instead of utilizing the Skorokhod representation theorem as in Huang (1992), Chapter 2, we apply a general Delta method to prove asymptotic normality. Observe that $\Lambda_U(x,y) = x + y - \lim_{t\to\infty} t(1 - C(1 - x/t, 1 - y/t))$. Thus, an estimator for $\Lambda_U(x,y)$ is given by

$$\begin{aligned} \hat{\Lambda}^{EVT}_{U,m}(x,y) &:= x + y - \frac{m}{k}\Big(1 - C_m\Big(1 - \frac{kx}{m}, 1 - \frac{ky}{m}\Big)\Big) \\ &\approx x + y - \frac{1}{k}\sum_{j=1}^{m} \mathbf{1}_{\{R^{(j)}_{m1} > m - kx \text{ or } R^{(j)}_{m2} > m - ky\}} \end{aligned} \tag{4.12}$$

with $k = k(m) \to \infty$ and $k/m \to 0$ as $m \to \infty$, where $\hat{\Lambda}^{EVT}_{L,m}$ is defined similarly. One important practical problem arises in the optimal choice of the parameter k which relates to the usual variance-bias problem. Some methods of choosing an optimal k are described below.

The main purpose of the first part of the present chapter concerns the study of the asymptotic behavior of the empirical tail-copulae $\hat{\Lambda}_{U,m}$ and $\hat{\Lambda}_{L,m}$ stated in (4.9) and (4.10). These estimators, although different, are related to the estimator $\hat{\Lambda}^{EVT}_{U,m}$ introduced in Huang (1992). However, as the method of proof used in Huang (1992) cannot be applied in our case, we chose a different approach. Furthermore, the results of Huang (1992) can be shown with the same techniques. Extensions of the latter results to dimensions larger than two are possible. Strong consistency and asymptotic normality of the above estimators are addressed in Sections 4.5 and 4.6.

Based on the above estimators, we propose

$$\hat{\lambda}_{U,m} := \hat{\Lambda}_{U,m}(1,1) \quad \text{and} \quad \hat{\lambda}^{EVT}_{U,m} := \hat{\Lambda}^{EVT}_{U,m}(1,1) \tag{4.13}$$

as nonparametric estimators for the upper tail-dependence coefficient and

$$\hat{\lambda}_{L,m} := \hat{\Lambda}_{L,m}(1,1) \quad \text{and} \quad \hat{\lambda}^{EVT}_{L,m} := \hat{\Lambda}^{EVT}_{L,m}(1,1) \tag{4.14}$$

as nonparametric estimators for the lower tail-dependence coefficient.

In Section 4.9.3, a simulation study compares both types of estimators for the tail-dependence coefficient regarding their finite sample properties. See also Section 4.9 for related results in the setting of elliptically contoured distributions where a parametric estimator is introduced.

4.3 Tail-copula properties

The name *tail copula* is justified by the results of the present section. Many properties of the tail copula are closely related to copula properties (cf. Nelsen (1999), Chapter 2).

Theorem 4.3.1 *If the limit functions $\Lambda_U(x,y)$ and $\Lambda_L(x,y)$, $(x,y)' \in \bar{\mathbb{R}}^2_+$, exist, they have the following properties.*

i) **(Groundedness)** $\Lambda_U(x,0) = \Lambda_U(0,y) = \Lambda_L(x,0) = \Lambda_L(0,y) = 0$ *for all* $x, y \in \bar{\mathbb{R}}_+$, *and* $\Lambda_U(x,\infty) = \Lambda_L(x,\infty) = x$ *and* $\Lambda_U(\infty,y) = \Lambda_L(\infty,y) = y$ *for all* $x, y \in \mathbb{R}_+$.

ii) **(Homogeneity)** $\Lambda_U(tx,ty) = t\Lambda_U(x,y)$ *and* $\Lambda_L(tx,ty) = t\Lambda_U(x,y)$ *for all* $t > 0$ *and* $(x,y)' \in \bar{\mathbb{R}}^2_+$.

iii) In the case of independence: $\Lambda_U(x,y) = \Lambda_L(x,y) = 0$ *for all* $(x,y)' \in \mathbb{R}^2_+$.

iv) **(Monotonicity)** $\Lambda_U(x,y)$ *and* $\Lambda_L(x,y)$ *are nondecreasing and Lipschitz continuous.*

v) $\Lambda_U(x,y)$ *and* $\Lambda_L(x,y)$ *are nonzero everywhere if they do not vanish in a single point* $(x,y)' \in \mathbb{R}^2_+$. *Hence* $\Lambda_U(x,y) = 0$ $(\Lambda_L(x,y) = 0)$ *for all* $(x,y)' \in \mathbb{R}^2_+$ *in case of upper (lower) tail-independence.*

vi) **(Uniformity)** *The limit relations for* $\Lambda_U(x,y)$ *and* $\Lambda_L(x,y)$ *are locally uniform in* $(x,y)' \in \mathbb{R}^2_+$.

Proof. Properties i), ii), and iii) follow immediately from Definition 2.2.7. Note that the limit of a regular varying function with index -1 is homogeneous.

iv) Consider e.g. $\Lambda := \Lambda_L$ and let C denote the corresponding copula. As the limit of nondecreasing functions, Λ_L is nondecreasing. Further, for $(x,y)', (\bar{x},\bar{y})' \in \bar{\mathbb{R}}^2_+$ we have

$$
\begin{aligned}
|\Lambda(x,y) - \Lambda(\bar{x},\bar{y})| &= \lim_{t\to\infty} t|C(x/t,y/t) - C(\bar{x}/t,\bar{y}/t)| \\
&\le |x-\bar{x}| + |y-\bar{y}| \le K\left|\left|\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}\right|\right|_2 \qquad (4.15)
\end{aligned}
$$

for some constant $K > 0$ because C is a bivariate distribution function with uniform margins.

v) The following inequalities hold for a, $b > 0$

$$
\min\{a,\, b\}\Lambda(x,y) \le \Lambda(a\,x,\, b\,y) \le \max\{a,\, b\}\Lambda(x,\, y)\,.
$$

To verify this, note that in case $a \le b$, using $\tau = t/a$ we find

$$
\begin{aligned}
\Lambda(a\,x,\, b\,y) &= \lim_{t\to\infty} t\,C(ax/t,\, by/t) \\
&= \lim_{\tau\to\infty} a\,\tau\,C(x/\tau,\, (b/a)y/\tau) = a\,\Lambda(x,\, (b/a)\,y) \ge a\,\Lambda(x,\, y)
\end{aligned}
$$

and the upper inequality follows similarly. Notice that the latter inequality also implies homogeneity. Next, if $\Lambda(x_0,\, y_0) > 0$ for some $x_0,\, y_0 > 0$, then we get

$$\Lambda(x,\, y) \geq \min\{x/x_0,\, y/y_0\}\Lambda(x_0,\, y_0) > 0\,.$$

vi) Finally, uniform convergence is obtained from the fact that for $x_n \to x_0,\, y_n \to y_0$ and $t_n \to \infty$, putting $\tau_n = t_n / \min\{x_n/x_0,\, y_n/y_0\}$ and $\xi_n = t_n / \max\{x_n/x_0,\, y_n/y_0\}$ we have

$$\begin{aligned}\min\{x_n/x_0,\, y_n/y_0\}\tau_n\, C(x_0/\tau_n,\, y_0/\tau_n) &\leq t_n\, C(x_n/t_n,\, y_n/t_n)\\ &\leq \max\{x_n/x_0,\, y_n/y_0\}\xi_n\, C(x_0/\xi_n,\, y_0/\xi_n).\end{aligned}$$

This implies that $t_n\, \Lambda(x_n/t_n,\, y_n/t_n) \to \Lambda(x_0,\, y_0)$ as $t_n \to \infty$. □

The next properties are given for the lower tail-copula only. However, analogous properties hold also for the upper pendant.

Theorem 4.3.2 *Suppose the limit function $\Lambda_L(x,y)$, $(x,y)' \in \bar{\mathbb{R}}_+^2$, exists. Then for all $(x,y)', (\bar{x},\bar{y})' \in \bar{\mathbb{R}}_+^2$ such that $x \leq \bar{x}, y \leq \bar{y}$ the following properties hold.*

i) **("Fréchet-Hoeffding bounds")** $0 \leq \Lambda_L(x,y) \leq \min\{x,y\}$.

ii) *For $a,\, b > 0$:* $\min\{a,\, b\}\Lambda_L(x,y) \leq \Lambda_L(a\,x,\, b\,y) \leq \max\{a,\, b\}\Lambda_L(x,\, y)$.

iii) **(2-increasing)** $\Lambda_L(\bar{x},\bar{y}) - \Lambda_L(\bar{x},y) - \Lambda_L(x,\bar{y}) + \Lambda_L(x,y) \geq 0$.

iv) **(Strict monotonicity)** *For $\Lambda_L \not\equiv 0 : \Lambda_L(x,y) < \Lambda_L(\bar{x},\bar{y})$ if $x < \bar{x}$ and $y < \bar{y}$.*

Proof. i) The lower and upper bound arise from the Fréchet-Hoeffding bounds for copulae (cf. Nelsen (1999), Theorem 2.2.3), i.e., for every copula function C we have

$$\max(u+v-1,0) \leq C(u,v) \leq \min(u,v).$$

Part ii) follows from the proof of part v) in Theorem 4.3.1. Part iii) is deduced from the fact that every copula is 2-increasing. Finally, the last part follows via part ii). □

Theorem 4.3.3 *Suppose the limit function $\Lambda_L(x,y)$, $(x,y)' \in \bar{\mathbb{R}}_+^2$, exists. Then, for any $y \in \bar{R}_+$ the derivative $\partial\Lambda_L/\partial x$ exists for almost all $x \in \mathbb{R}_+$, and for such x and y*

$$0 \leq \frac{\partial}{\partial x}\Lambda_L(x,y) \leq 1. \tag{4.16}$$

Similarly, for any $x \in \mathbb{R}_+$ the partial derivative $\partial\Lambda_L/\partial y$ exists for almost all $y \in \mathbb{R}_+$, and for such x and y

$$0 \leq \frac{\partial}{\partial y}\Lambda_L(x,y) \leq 1. \tag{4.17}$$

Furthermore, the functions $x \mapsto \partial\Lambda_L(x,y)/\partial y$ and $y \mapsto \partial\Lambda_L(x,y)/\partial x$ are defined and nondecreasing almost everywhere on $\bar{\mathbb{R}}_+$.

Proof. The partial derivatives $\partial\Lambda_L/\partial x$ and $\partial\Lambda_L/\partial y$ exist because monotone functions are differentiable almost everywhere (c.f Theorem 7.2.1 in Wheeden and Zygmund (1977)). Inequalities (4.16) and (4.17) follow from the Lipschitz condition (4.15). Further, for fixed $y \leq \bar{y}$ the function $y \mapsto \Lambda_L(x,y) - \Lambda_L(x,\bar{y})$ is nondecreasing according to part iii) in Theorem 4.3.2. Thus $\partial(\Lambda_L(x,y) - \Lambda_L(x,\bar{y}))/\partial x$ is defined and nonnegative almost everywhere. The final assertion is now immediate. □

Remark. The above properties can be adjusted if the tail copula does not exist on the entire space $\bar{\mathbb{R}}^2_+$.

4.4 The space $\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$ and the Delta method

In the present section, the function space $\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$ is defined and weak convergence is established. This space turns out to be appropriate in the context of empirical tail-copulae. Further, we formulate a general *Delta method* which is utilized to prove asymptotic normality for empirical tail-copulae. We equip the space with a uniform metric on compacta and establish a necessary and sufficient condition for weak convergence.

Consider the metric spaces $(\mathbb{D}, d)$ and $(\mathbb{E}, e)$. Concepts of weak convergence and almost-sure convergence are traditionally applied to Borel probability measures defined on some space $(\mathbb{D}, \mathcal{D})$ with $\mathcal{D}$ denoting the Borel σ-field of $\mathbb{D}$; see for example Billingsley (1968), p. 68. In particular, $\mathcal{D}$ is the smallest σ-field generated by the open sets. However, in the context of empirical tail-copulae living in some space $(\mathbb{D}, \mathcal{D})$, these concepts must be modified as no probability measures can be defined on the corresponding Borel σ-field $\mathcal{D}$. Loosely speaking, the Borel σ-field turns out to be too large. To overcome this obstacle, several approaches can be distinguished. First, we could restrict to a smaller σ-field like the ball σ-field $\mathcal{D}_B$, and define weak convergence on the new space $(\mathbb{D}, \mathcal{D}_B)$; see for instance Dudley (1966), Dudley (1967) and Pollard (1984). Second, the metric d could be adjusted in such a way that the classical theory is still applicable. A famous example is the Skorokhod metric on the càdlàg space $D[0,1]$; see Skorokhod (1956). However, we will utilize the concepts of weak convergence and almost-sure convergence defined by outer expectations. A good reference for this theory is the book by Van der Vaart and Wellner (1996).

Definition 4.4.1 (Weak convergence with respect to outer expectations)
Let Y be an arbitrary (not necessarily measurable) map from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to the extended real line $\bar{\mathbb{R}}$. The outer integral of Y with respect to the probability measure $\mathbb{P}$ is defined as

$$\mathbb{E}^*Y = \inf\{\mathbb{E}U : U \geq Y, U : \Omega \to \bar{\mathbb{R}} \text{ measurable and } \mathbb{E}U \text{ exists}\}.$$

For each $n \geq 1$, let X_n be an arbitrary (not necessarily measurable) map from a probability space $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ to a metric space $(\mathbb{D}, d)$. Then, X_n is said to converge weakly ($\xrightarrow{w}$) to a Borel-measurable map X, if

$$\mathbb{E}^* f(X_n) \to \mathbb{E} f(X) \text{ for every } f \in C_b(\mathbb{D}),$$

where $C_b(I\!\!D)$ denotes the set of all bounded, continuous and real functions on $I\!\!D$.

In order to define the space $\mathcal{B}_\infty(\bar{I\!\!R}^2_+)$ together with an appropriate metric we need some more notation: The space $l^\infty(T)$ for an arbitrary set T is defined as the set of all uniformly bounded, real functions on T, precisely all functions $f : T \to I\!\!R$ such that

$$||f||_T := \sup_{t \in T} |f(t)| < \infty.$$

Consequently, the uniform distance on $l^\infty(T)$ is defined by

$$d(f_1, f_2) = ||f_1 - f_2||_T.$$

The stochastic processes $\{X_n(t) : t \in T\}$ considered below will have their sample paths in $l^\infty(T)$ if T is a compact subset of $\bar{I\!\!R}^2_+$. If not stated otherwise, T always denotes a compact subset of $\bar{I\!\!R}^2_+$ in the following.

The main advantage of Definition 4.4.1 arises from its applicability to general empirical processes as classical theorems like the continuous mapping theorem and Prohorov's theorem can be established in the new setting; see Van der Vaart and Wellner (1996). Once the latter theorems are established, the convergence theory becomes less technical like for instance a multidimensional Skorokhod construction (see Neuhaus (1971)). The space $l^\infty(T)$ is equipped with the Borel σ-field. The only measurability property we require for X_n is that the maps $X_n(t) : \Omega_n \to I\!\!R$ are measurable (this means that $X_n(t)$ is a random variable for each fixed $t \in T$, cf. Pollard (1984)); a rather weak condition. However, the limiting process X which turns out to be a Brownian bridge is a Borel measurable map $X|_T : \Omega \to C(T)$ as the space $C(T)$ of all continuous real function is a separable and complete subspace of $l^\infty(T)$ under its uniform metric. Moreover, it can be shown that the Borel σ-field of $C(T)$ correspond to the projection σ-field.

We are ready to define the metric space $\mathcal{B}_\infty(\bar{I\!\!R}^2_+)$.

Definition 4.4.2 *The space $\mathcal{B}_\infty(\bar{I\!\!R}^2_+)$ is defined as the family of all functions $f : \bar{I\!\!R}^2_+ \to I\!\!R$ which are locally uniformly-bounded on every compact subset of $\bar{I\!\!R}^2_+$ (but not necessarily on $\bar{I\!\!R}^2_+$). Then, $\mathcal{B}_\infty(\bar{I\!\!R}^2_+)$ is a complete metric space under the metric*

$$d(f_1, f_2) = \sum_{i=1}^{\infty} 2^{-i}(||f_1 - f_2||_{T_i} \wedge 1) \tag{4.18}$$

with $T_{3i} = T_{3i-1} \cup [0, i]^2$, $T_{3i-1} = T_{3i-2} \cup ([0, i] \times \{\infty\})$, $T_{3i-2} = T_{3(i-1)} \cup (\{\infty\} \times [0, i]) \subset \bar{I\!\!R}^2_+$, $i \in I\!\!N$, and $T_0 = \emptyset$. Thus a sequence of elements in $\mathcal{B}_\infty(\bar{I\!\!R}^2_+)$ converges in this metric if it converges uniformly on each T_i.

Remark. Later we need the space $\mathcal{B}_\infty(\bar{I\!\!R}^n_+)$ which is defined analogously to $\mathcal{B}_\infty(\bar{I\!\!R}^2_+)$.

The following theorem is fundamental for our purposes. A *proof* can be found in Van der Vaart and Wellner (1996), Theorem 1.6.1.

Theorem 4.4.3 *For each $n \geq 1$, let $X_n : \Omega_n \to \mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$ be an arbitrary map. Then the sequence X_n converges weakly to a tight limit if and only if every sequence of restrictions $X_n|_{T_i} : \Omega_n \to l^\infty(T_i)$ converges weakly to a tight limit.*

The Delta method (see Casella and Berger (2002), Section 5.5.4) is a well-known technique in statistics to prove results concerning asymptotic normality of an estimator. In the context of tail copulae we need a quite general version of the Delta method. For this, the notion of Hadamard differentiability is needed. Let $(\mathbb{D}, d)$ and $(\mathbb{E}, e)$ be metrizable and topological vector spaces, in particular vector addition and scalar multiplication are continuous operations.

Definition 4.4.4 (Hadamard differentiability) *A map $\phi : \mathbb{D}_\phi \subset \mathbb{D} \to \mathbb{E}$ is called Hadamard-differentiable at $\theta \in \mathbb{D}_\phi$ if there exists a continuous linear map $\phi'_\theta : \mathbb{D} \to \mathbb{E}$ such that*

$$\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} \to \phi'_\theta(h), \quad \text{as } n \to \infty, \tag{4.19}$$

for all converging sequences $t_n \to 0$ and $h_n \to h$ such that $\theta + t_n h_n \in \mathbb{D}_\phi$ for all n. Further, $\phi : \mathbb{D}_\phi \subset \mathbb{D} \to \mathbb{E}$ is called Hadamard-differentiable tangentially to a set $\mathbb{D}_0 \subset \mathbb{D}$ by requiring that $h_n \to h$ with $h \in \mathbb{D}_0$. In that case the derivative ϕ'_θ needs only to be defined on $\mathbb{D}_0$.

Note that $\mathbb{D}_\phi$ is allowed to be any arbitrary subset of $\mathbb{D}$; this turns out to be important later.

Theorem 4.4.5 (Delta method) *Let $\phi : \mathbb{D}_\phi \subset \mathbb{D} \to \mathbb{E}$ be Hadamard-differentiable at θ tangentially to $\mathbb{D}_0$. Suppose $X_n : \Omega_n \to \mathbb{D}_\phi$ are (not necessarily measurable) maps with $r_n(X_n - \theta) \xrightarrow{w} X$ for some sequence of constants $r_n \to \infty$, where $X : \Omega \to \mathbb{D}_\phi$ is separable. Then*

$$r_n\big(\phi(X_n) - \phi(\theta)\big) \xrightarrow{w} \phi'_\theta(X).$$

For details we refer the reader to Van der Vaart and Wellner (1996), p. 374. We are ready to prove asymptotic normality and strong consistency of the proposed estimators.

4.5 Asymptotic normality

The proof of asymptotic normality of the estimators $\hat{\Lambda}_{U,m}(x, y)$ and $\hat{\Lambda}_{L,m}(x, y)$ is accomplished in two steps. In the first step we assume that the margins G and H are known, and provide the asymptotic normality. In the second step we assume that the marginal distribution functions G and H are unknown, and prove the asymptotic result by utilizing the Delta method (see Theorem 4.4.5). In the case of known marginal distribution functions G and H we consider the following estimator for $\Lambda_{U,m}(x, y)$ and $\Lambda_{L,m}(x, y)$:

$$\hat{\Lambda}^*_{U,m}(x, y) := \frac{1}{k} \sum_{j=1}^{m} \mathbf{1}_{\{G(X^{(j)}) > 1 - \frac{kx}{m} \text{ and } H(Y^{(j)}) > 1 - \frac{ky}{m}\}} \tag{4.20}$$

and

$$\hat{\Lambda}^*_{L,m}(x,y) := \frac{1}{k}\sum_{j=1}^{m} \mathbf{1}_{\left\{G(X^{(j)}) \le \frac{kx}{m} \text{ and } H(Y^{(j)}) \le \frac{ky}{m}\right\}}. \tag{4.21}$$

Condition 4.5.1 (Second order condition) *The lower tail-copula $\Lambda_L(x,y)$ is said to satisfy a second order condition if a function $A : \mathbb{R}_+ \to \mathbb{R}_+$ exists such that A is regularly varying, $A(t) \to 0$ as $t \to \infty$, and*

$$\lim_{t\to\infty} \frac{\Lambda_L(x,y) - tC(x/t, y/t)}{A(t)} = g(x,y) < \infty$$

locally uniformly for $(x,y)' \in \bar{\mathbb{R}}^2_+$ and some nonconstant function g. The second order condition for the upper tail-copula is defined analogously.

Note that $A(t)$ is regularly varying at infinity so this is just a second order condition on regular variation (cf. de Haan and Stadtmüller (1996)).

Theorem 4.5.2 (Asymptotic normality under known margins G and H) *Let F be a bivariate distribution function with continuous marginal distribution functions G and H. Suppose the tail copulae $\Lambda_U \not\equiv 0$ and $\Lambda_L \not\equiv 0$ exist and the Second order condition 4.5.1 with*

$$\sqrt{k}\, A(m/k) \to 0 \quad as \quad m \to \infty \tag{4.22}$$

holds for some sequence $k = k(m) \to \infty$ and $k/m \to 0$. Then

$$\sqrt{k}\left(\hat{\Lambda}^*_{U,m}(x,y) - \Lambda_U(x,y)\right) \xrightarrow{w} \mathbb{G}_{\hat{\Lambda}^*_U}(x,y) \tag{4.23}$$

and

$$\sqrt{k}\left(\hat{\Lambda}^*_{L,m}(x,y) - \Lambda_L(x,y)\right) \xrightarrow{w} \mathbb{G}_{\hat{\Lambda}^*_L}(x,y), \tag{4.24}$$

where $\mathbb{G}_{\hat{\Lambda}^*_U}(x,y)$ and $\mathbb{G}_{\hat{\Lambda}^*_L}(x,y)$ are centered tight continuous Gaussian random fields, weak convergence takes place in $\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$, and the covariance structure of $\mathbb{G}_{\hat{\Lambda}^*_U}(x,y)$ and $\mathbb{G}_{\hat{\Lambda}^*_L}(x,y)$ is given in Corollary 4.5.4 below.

Proof. The proof of asymptotic normality is given for the lower empirical tail-copula $\hat{\Lambda}^*_{L,m}(x,y)$. The upper empirical tail-copula is similarly treated. First we prove that

$$\alpha_m(x,y) := \sqrt{k}\left(\hat{\Lambda}^*_{L,m}(x,y) - \frac{m}{k}C\Big(\frac{kx}{m}, \frac{ky}{m}\Big)\right) \xrightarrow{w} \mathbb{G}_{\hat{\Lambda}^*_L}(x,y) \quad \text{as } m \to \infty$$

with $k = k(m) \to \infty$ and $k/m \to 0$ as $m \to \infty$ and a centered tight Gaussian process $\mathbb{G}_{\hat{\Lambda}^*_L}(x,y)$. Further, weak convergence takes place in $\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$. Therefore, finite dimensional convergence and tightness have to be established.

i) (Finite dimensional convergence) We have to show that the finite-dimensional projections of $\alpha_m(x,y)$ converge in distribution to a normal random vector, i.e., for each finite subset $\{(x_1,y_1),\dots,(x_t,y_t)\}$ of $\bar{\mathbb{R}}_+^2$ there exists a centered normal random vector $(\alpha(x_1,y_1),\dots,\alpha(x_t,y_t))$ with appropriate covariance structure such that

$$(\alpha_m(x_1,y_1),\dots,\alpha_m(x_t,y_t)) \xrightarrow{d} (\alpha(x_1,y_1),\dots,\alpha(x_t,y_t)).$$

The latter is shown by a multivariate version of the Lindeberg-Feller theorem for triangular arrays (see Durrett (1996), p.116, or Araujo and Giné (1980), p.41). Let $\{(x_1,y_1),\dots,(x_t,y_t)\}$ be an arbitrary but fixed finite subset of $\bar{\mathbb{R}}_+^2$. Put

$$Z_{i,m}^{(j)} := \frac{1}{\sqrt{k}}\mathbf{1}_{\{G(X^{(j)}) \le \frac{kx_i}{m} \text{ and } H(Y^{(j)}) \le \frac{ky_i}{m}\}} - \frac{1}{\sqrt{k}}C(\frac{kx_i}{m},\frac{ky_i}{m})$$

for all $i=1,\dots,t$. Then $\mathbb{E}(Z_{i,m}^{(j)})=0$ for all $i=1,\dots,t$. For every $r,s\in\{1,\dots,t\}$

$$\begin{aligned}
&\sum_{j=1}^m \mathbb{E}(Z_{r,m}^{(j)}Z_{s,m}^{(j)})\\
&\quad= \frac{m}{k}\left\{\mathbb{P}\Big(G(X^{(j)}) \le \frac{k}{m}\min\{x_r,x_s\} \text{ and } H(Y^{(j)}) \le \frac{k}{m}\min\{y_r,y_s\}\Big)\right.\\
&\qquad\left. -C\left(\frac{kx_r}{m},\frac{ky_r}{m}\right)C\left(\frac{kx_s}{m},\frac{ky_s}{m}\right)\right\}\\
&\quad\to \Lambda_L(\min\{x_r,x_s\},\min\{y_r,y_s\}) =: a_{r,s} \text{ as } m\to\infty.
\end{aligned}$$

Notice that $tC(x_r/t,y_r/t)C(x_s/t,y_s/t)\to 0$ as $t\to\infty$. The matrix $A=(a_{r,s})_{r,s=1,\dots,t}$ is nonzero if $\Lambda_L \not\equiv 0$ according to Theorem 4.3.1. Further, for $Z_m^{(j)}=(Z_{1,m}^{(j)},\dots,Z_{t,m}^{(j)})$ and the Euclidian norm $||.||_2$ we have

$$||Z_m^{(j)}||_2^2 = \sum_{i=1}^t (Z_{i,m}^{(j)})^2 \le \frac{t}{k}$$

and thus

$$\sum_{j=1}^m \int_{\{||Z_m^{(j)}||_2>\varepsilon\}} ||Z_m^{(j)}||_2^2 d\mathbb{P} \to 0 \quad \text{as } m\to\infty$$

for every $\varepsilon>0$. Therefore

$$(\alpha_m(x_1,y_1),\dots,\alpha_m(x_t,y_t)) \xrightarrow{d} (\alpha(x_1,y_1),\dots,\alpha(x_t,y_t)) \sim N(0,A)$$

with $A=(a_{r,s})$.

ii) (Tightness) First we prove tightness on $[0,M]^2$ for every fixed $M\in\mathbb{N}$ via asymptotic uniform equicontinuity in probability of α_m, i.e, for each $\xi>0$ and $\eta>0$ there exist $\delta\in(0,1)$ and $m_0\in\mathbb{N}$ such that

$$\mathbb{P}\Big(\sup_{\substack{|x_1-x_2|^2+|y_1-y_2|^2<\delta\\ x_i,y_i\in[0,M],\, i=1,2}} |\alpha_m(x_1,y_1)-\alpha_m(x_2,y_2)| \ge \xi\Big) \le \eta \quad \forall m\ge m_0.$$

Note that α_m belongs to the space of cádlág functions $D(\bar{I\!R}_+^2)$. It can be shown that the ball-σ-field coincides with the projection σ-field in the space $D(\bar{I\!R}_+^2)$ and therefore α_m is measurable with respect to the ball-σ-field. This justifies to take probability instead of outer probability in the above expression. Tightness is now shown by the following reasoning. Consider a partition of $[0,M]^2$ into equally sized cubes $I_{l,L}$ with partition points $(M\cdot l_1/L, M\cdot l_2/L)$, $l_i\in\{0,\dots,L\}, L\in I\!N$, $i=1,2$. Then for arbitrary but fixed $\xi>0$ and $\delta\in(0,1)$ such that $1/L\geq\delta$ we obtain

$$\begin{aligned}
&I\!P\Big(\sup_{\substack{|x_1-x_2|^2+|y_1-y_2|^2<\delta\\ x_i,y_i\in[0,M],\, i=1,2}}|\alpha_m(x_1,y_1)-\alpha_m(x_2,y_2)|\geq\xi\Big)\\
&\quad\leq I\!P(3\max_{\substack{0\leq l_i<L\\ i=1,2}}\sup_{\binom{x_1}{y_1},\binom{x_2}{y_2}\in I_{l,L}}|\alpha_m(x_1,y_1)-\alpha_m(x_2,y_2)|\geq\xi)=:I_1.
\end{aligned}$$

Without loss of generality we assume $x_1<x_2$ and $y_1<y_2$. Then,

$$\begin{aligned}
I_1\leq&\sum_{\substack{0\leq l_i<L\\ i=1,2}}\Big\{I\!P\Big(\sup_{\binom{x_1}{y_1},\binom{x_2}{y_2}\in I_{l,L}}\sqrt{m}\Big|\frac{1}{m}\sum_{j=1}^m\mathbf{1}_{\{\frac{kx_1}{m}<G(X^{(j)})\leq\frac{kx_2}{m}\text{ and }H(Y^{(j)})\leq\frac{ky_2}{m}\}}\\
&\qquad-C\Big(\Big(\frac{kx_1}{m},\frac{kx_2}{m}\Big]\times\Big[0,\frac{ky_2}{m}\Big]\Big)\Big|\geq\frac{1}{2}\frac{\sqrt{k}}{\sqrt{m}}\frac{\xi}{3}\Big)\\
&\qquad+I\!P\Big(\sup_{\binom{x_1}{y_1},\binom{x_2}{y_2}\in I_{l,L}}\sqrt{m}\Big|\frac{1}{m}\sum_{j=1}^m\mathbf{1}_{\{G(X^{(j)})\leq\frac{kx_1}{m}\text{ and }\frac{ky_1}{m}<H(Y^{(j)})\leq\frac{ky_2}{m}\}}\\
&\qquad-C\Big(\Big[0,\frac{kx_1}{m}\Big]\times\Big(\frac{ky_1}{m},\frac{ky_2}{m}\Big]\Big)\Big|\geq\frac{1}{2}\frac{\sqrt{k}}{\sqrt{m}}\frac{\xi}{3}\Big)\Big\}\\
\leq&\sum_{n=1}^2\sum_{\substack{0\leq l_i<L\\ i=1,2}}c\cdot\exp\left(-\frac{\eta^2k\xi^2}{36mC(A_{n,m}^{l_i,L})}\cdot\psi\Big(\frac{\sqrt{k}\eta}{mC(A_{n,m}^{l_i,L})}\Big)\right)=:I_2,
\end{aligned}$$

where the constants $c,\eta>0$ are independent of the other parameters, and

$$\begin{aligned}
A_{1,m}^{l_i,L}&:=\Big(M\frac{kl_1}{mL},M\frac{k(l_1+1)}{mL}\Big]\times\Big[0,M\frac{k(l_2+1)}{mL}\Big],\text{ and}\\
A_{2,m}^{l_i,L}&:=\Big[0,M\frac{kl_1}{mL}\Big]\times\Big(M\frac{kl_2}{mL},M\frac{k(l_2+1)}{mL}\Big].
\end{aligned}$$

The last inequality is due to Ruymgaart and Wellner (1982), inequality 1.1. In particular the function $\psi:[-1,\infty)\to I\!R$ satisfies $\psi(0)=1$, $\psi(x)\sim(2\log x)/x\to0$ as $x\to\infty$, ψ is decreasing and continuous, and $(\cdot)\psi(\cdot)$ is increasing. Observe that $C(A_{n,m}^{l_i,L})\leq\frac{k}{Lm}$ for all $l_i\in\{0,\dots,L\}$, $i=1,2$.

Distinguish two cases: Either for $m,L\in I\!N$, $n=1,2$, and $l_i\in\{0,\dots,L\}$, $i=1,2$,

$$\frac{\sqrt{k}\eta}{mC(A_{n,m}^{l_i,L})}\leq1\quad\text{or}\quad\frac{\sqrt{k}\eta}{mC(A_{n,m}^{l_i,L})}>1.$$

In the first case an upper bound is provided by

$$I_2 \le 2L^2 c \cdot \exp\Big(-\frac{\eta^2\xi^2}{36} L\psi(1)\Big), \quad L \in \mathbb{N},$$

whereas in the second case we utilize the upper bound

$$I_2 \le 2L^2 c \cdot \exp\Big(-\frac{\eta\xi^2}{36} \sqrt{k}\psi(1)\Big) \to 0 \quad \text{as } m \to \infty.$$

This immediately yields tightness on $[0, M]^2$ for every fixed $M \in \mathbb{N}$. Tightness on $[0, M] \times \{\infty\}$ and $\{\infty\} \times [0, M]$ is shown along the same lines.

iii) According to part ii), Theorem 1.5.7 and Lemma 1.3.8 in Van der Vaart and Wellner (1996) the sequence of restrictions $\alpha_m|_{T_i}$ with T_i as defined in Definition 4.4.2 is asymptotical tight. This follows because the limiting process $\mathbb{G}_{\hat\Lambda_L^*}(x, y)|_{T_i}$ is tight in $\mathbb{R}$ for every $(x, y)' \in T_i$ as its law is a Borel probability-measure on a Polish space. Hence, the sequence of restrictions $\alpha_m|_{T_i}$ weakly converges to the tight limit $\mathbb{G}_{\hat\Lambda_L^*}(x, y)|_{T_i}$ due to part i) and Theorem 1.5.4 in Van der Vaart and Wellner (1996). Finally, weak convergence of α_m in $\mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)$ is provided by Theorem 4.4.3. Continuity of the sample paths in $\bar{\mathbb{R}}_+^2$ follows according to the Addendum 1.5.8 in the latter reference.

iv) (Second order condition) To prove the desired weak convergence result

$$\sqrt{k}\big\{\hat\Lambda_{L,m}^*(x, y) - \Lambda_L(x, y)\big\} \xrightarrow{w} \mathbb{G}_{\hat\Lambda_L^*}(x, y)$$

it remains to show that

$$\sqrt{k}\Big\{\Lambda_L(x, y) - \frac{m}{k} C\Big(\frac{kx}{m}, \frac{ky}{m}\Big)\Big\} \to 0 \quad \text{as } m \to \infty$$

uniformly for $(x, y)' \in \bar{\mathbb{R}}_+^2$. This convergence is implied by the Second order condition 4.5.1 and the fact that $\sqrt{k}A(m/k) \to 0$ as $m \to \infty$. □

Remark. If the tail copula is only defined on some subinterval of $\bar{\mathbb{R}}_+^2$, the latter results hold on this subinterval of $\bar{\mathbb{R}}_+^2$ only.

Theorem 4.5.3 (Asymptotic normality under unknown margins G and H)
Let F be a bivariate distribution function with continuous marginal distribution functions G and H. If the tail copulae $\Lambda_U \not\equiv 0$ and $\Lambda_L \not\equiv 0$ exist, possess continuous partial derivatives, and the Second order condition 4.5.1 *holds, then for $\sqrt{k}A(m/k) \to 0$ as $m \to \infty$*

$$\sqrt{k}\big\{\hat\Lambda_{U,m}(x, y) - \Lambda_U(x, y)\big\} \xrightarrow{w} \mathbb{G}_{\Lambda_U}(x, y),$$

and

$$\sqrt{k}\big\{\hat\Lambda_{L,m}(x, y) - \Lambda_L(x, y)\big\} \xrightarrow{w} \mathbb{G}_{\Lambda_L}(x, y),$$

where $\mathbb{G}_{\Lambda_U}(x, y)$ and $\mathbb{G}_{\Lambda_L}(x, y)$ are centered tight continuous Gaussian random fields, weak convergence takes place in $\mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)$, and the covariance structure of $\mathbb{G}_{\hat\Lambda_U}(x, y)$ and $\mathbb{G}_{\hat\Lambda_L}(x, y)$ is given in Corollary 4.5.4 *below.*

Proof. The proof is only given for the lower tail-copula as the upper tail-copula can be treated similarly. The space of locally uniformly bounded real function on compact sets of $\mathbb{R}_+$ is denoted by $\mathcal{B}_\infty(\mathbb{R}_+)$; the corresponding metric is defined analogously to (4.18). Let $B^I(\mathbb{R}_+) \subset \mathcal{B}_\infty(\mathbb{R}_+)$ denote the set of all nondecreasing functions $\zeta : \mathbb{R}_+ \mapsto \mathbb{R}_+$. Define the set

$$\mathcal{B}^I_\infty(\bar{\mathbb{R}}^2_+) := \{\gamma \in \mathcal{B}_\infty(\bar{\mathbb{R}}^2_+) \mid \gamma(\cdot,\infty) \in B^I(\mathbb{R}_+) \text{ and } \gamma(\infty,\cdot) \in B^I(\mathbb{R}_+)\}.$$

We apply the Delta method introduced in Section 4.4 to the following map

$$\Phi : \mathcal{B}^I_\infty(\bar{\mathbb{R}}^2_+) \mapsto \mathcal{B}_\infty(\bar{\mathbb{R}}^2_+).$$

For the definition of Φ we need some additional notation: Let ζ^- denote the adjusted generalized inverse function of $\zeta \in B^I(\mathbb{R}_+)$ defined by

$$\zeta^-(p) := \begin{cases} \zeta^{-1}(p) & \text{if } \zeta^{-1}(p) < \infty, \\ \lim_{z\to\infty} \zeta(z) & \text{if } \zeta^{-1}(p) = \infty, \end{cases}$$

where ζ^{-1} refers to the generalized inverse function given by (2.7). Split the set $\bar{\mathbb{R}}^2_+$ into three subsets $S_1 := \mathbb{R}^2_+$, $S_2 := [0,\infty) \times \{\infty\}$, and $S_3 := \{\infty\} \times [0,\infty)$. For some arbitrary function $\gamma \in \mathcal{B}^I_\infty(\bar{\mathbb{R}}^2_+)$ the map Φ is defined for $(x,y)' \in S_1$ by

$$\Phi : \gamma(x,y) \overset{\Phi_1}{\mapsto} (\gamma(x,y), \gamma(x,\infty), \gamma(\infty,y))$$

$$\overset{\Phi_2}{\mapsto} (\gamma(x,y), \gamma^-(x,\infty), \gamma^-(\infty,y)) \overset{\Phi_3}{\mapsto} \gamma \circ (\gamma^-(x,\infty), \gamma^-(\infty,y)),$$

for $(x,y)' \in S_2$ by

$$\Phi : \gamma(x,y) \overset{\Phi_1}{\mapsto} (\gamma(x,y), \gamma(x,\infty), \gamma(x,\infty))$$

$$\overset{\Phi_2}{\mapsto} (\gamma(x,y), \gamma^-(x,\infty), \gamma^-(x,\infty)) \overset{\Phi_3}{\mapsto} \gamma \circ (\gamma^-(x,\infty), \infty),$$

and for $(x,y)' \in S_3$ by

$$\Phi : \gamma(x,y) \overset{\Phi_1}{\mapsto} (\gamma(x,y), \gamma(\infty,y), \gamma(\infty,y))$$

$$\overset{\Phi_2}{\mapsto} (\gamma(x,y), \gamma^-(\infty,y), \gamma^-(\infty,y)) \overset{\Phi_3}{\mapsto} \gamma \circ (\infty, \gamma^-(\infty,y)).$$

The spaces $C(\bar{\mathbb{R}}^2_+) \subset \mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$ and $C(\mathbb{R}_+) \subset \mathcal{B}_\infty(\mathbb{R}_+)$ consist of all continuous functions in $\mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$ and $\mathcal{B}_\infty(\mathbb{R}_+)$, respectively. In order to apply the Delta method we have to show that the map Φ is Hadamard-differentiable on $\mathcal{B}^I_\infty(\bar{\mathbb{R}}^2_+)$ at $\gamma_0 = \Lambda_L$ tangentially to $C(\bar{\mathbb{R}}^2_+) \subset \mathcal{B}_\infty(\bar{\mathbb{R}}^2_+)$.

i) The first map Φ_1 is Hadamard-differentiable on $\mathcal{B}^I_\infty(\bar{\mathbb{R}}^2_+)$ at Λ_L tangentially to $C(\bar{\mathbb{R}}^2_+)$ as it is linear and continuous.

ii) The second map Φ_2 is Hadamard-differentiable on $\mathcal{B}^I_\infty(\bar{\mathbb{R}}^2_+) \times B^I(\mathbb{R}_+) \times B^I(\mathbb{R}_+)$ at $(\Lambda_L, \mathrm{id}_{\mathbb{R}_+}, \mathrm{id}_{\mathbb{R}_+})$ tangentially to $C(\bar{\mathbb{R}}^2_+) \times C(\mathbb{R}_+) \times C(\mathbb{R}_+)$. Note that Hadamard differentiability of Φ_2 is equivalent to Hadamard differentiability of the respective (vector) components of Φ_2. The first (vector) component of Φ_2 is Hadamard-differentiable as it represents the identity map on $\mathcal{B}_\infty^I(\bar{I\!R}_+^2)$. The second and third (vector) components are Hadamard-differentiable because the adjusted generalized inverse maps

$$\gamma(\cdot,\infty) \to \gamma^-(\cdot,\infty) \quad \text{and} \quad \gamma(\infty,\cdot) \to \gamma^-(\infty,\cdot)$$

are Hadamard-differentiable on $B^I(I\!R_+)$ at $\mathrm{id}_{I\!R_+}$ tangentially to $C(I\!R_+)$. Here

$$\gamma_0(\cdot,\infty) = \gamma_0(\infty,\cdot) = \Lambda_L(\cdot,\infty) = \Lambda_L(\infty,\cdot) = \mathrm{id}_{I\!R_+}$$

denote the identity function on R_+.

We restrict ourselves to the map $\gamma(\cdot,\infty) \to \gamma^-(\cdot,\infty)$. Set $\bar{\gamma}_0(\cdot) := \gamma_0(\cdot,\infty) = \mathrm{id}_{I\!R_+}$. Consider the sequence $h_t \to h$ as $t \to 0$ where $h \in C(I\!R_+)$ (and $h_t \in \mathcal{B}_\infty(I\!R_+)$), i.e., h is a continuous function on $I\!R_+$, such that $\bar{\gamma}_0 + th_t \in B^I(I\!R_+)$ for every t . Let $p \in I\!R_+$ then $\bar{\gamma}_0^-(p) = p$. Abbreviate $(\bar{\gamma}_0 + th_t)^-(p)$ to ξ_{pt} and notice that $\xi_{pt} \in I\!R_+$. Setting $\varepsilon_{pt} := t^2 \wedge \xi_{pt} \geq 0$ yields for $p \in [0, M_t)$ with $M_t := \lim_{z\to\infty}(\bar{\gamma}_0 + th_t)(z)$

$$(\bar{\gamma}_0 + th_t)(\xi_{pt} - \varepsilon_{pt}) \leq p \leq (\bar{\gamma}_0 + th_t)(\xi_{pt}).$$

Further $\bar{\gamma}_0(\xi_{pt}) = \xi_{pt}$ and $\bar{\gamma}_0(\xi_{pt} - \varepsilon_{pt}) = \xi_{pt} - \varepsilon_{pt}$ for all $p \in I\!R_+$. Thus it follows that

$$-th(\xi_{pt}) + o(t) \leq \xi_{pt} - p \leq -th(\xi_{pt} - \varepsilon_{pt}) + o(t), \tag{4.25}$$

where the $o(t)$-terms are uniform in $p \in [0, M_t)$. Note that $M_t \to \infty$ as $t \to 0$ because $h_t \to h$ with $h \in C(I\!R_+)$.

Finally, $h(\xi_{pt}) \to h(p)$ and $h(\xi_{pt} - \varepsilon_{pt}) \to h(p)$ uniformly in $p \in I\!R_+$ because h is continuous on R_+ and $\xi_{pt} \to p$ uniformly in $p \in I\!R_+$ on $\mathcal{B}_\infty(I\!R_+)$. The latter claim is proven if we show that $|\xi_{pt} - p| = O(t)$ uniformly in p on the interval $[0,T]$ for every arbitrary but fixed T. According to (4.25), it suffices to show that $h(\xi_{pt})$ is uniformly bounded on $[0,T]$. However, the function h is continuous on $I\!R_+$, therefore, we must prove that ξ_{pt} is uniformly bounded on $[0,T]$. This follows by the definition of the adjusted generalized inverse function and the fact that $h_t \to h$ as $t \to 0$ (h is continuous) on $\mathcal{B}_\infty(I\!R_+)$.

Hence Hadamard differentiability of $\gamma(\cdot,\infty) \to \gamma^-(\cdot,\infty)$ holds and its derivative is given by the linear map $h \mapsto -h$.

iii) The third map Φ_3 (composition map) is Hadamard-differentiable on $\mathcal{B}_\infty^I(\bar{I\!R}_+^2) \times B^I(I\!R_+) \times B^I(I\!R_+)$ at $(\Lambda_L, \mathrm{id}_{I\!R_+}, \mathrm{id}_{I\!R_+})$ tangentially to $C(\bar{I\!R}_+^2) \times C(I\!R_+) \times C(I\!R_+)$. according to Lemma A (in the appendix). Uniform Fréchet differentiability in Lemma A is implied by the continuous partial derivatives of Λ_L which yield (uniformly) continuous differentiability of Λ_L with respect to the norm (4.18) (cf. Heuser (2000) , Satz 164.4, and Van der Vaart and Wellner (1996), Problem 1, p. 397).

iv) Hadamard differentiability of Φ on $\mathcal{B}_\infty^I(\bar{I\!R}_+^2)$ at $\gamma_0 = \Lambda_L$ tangentially to $C(\bar{I\!R}_+^2) \subset \mathcal{B}_\infty(\bar{I\!R}_+^2)$ follows now with the chain rule (Lemma 3.9.3. in Van der Vaart and Wellner (1996)).

v) The final steps link the Delta method to the desired weak convergence result. Note that $\gamma_0(x,y) = \Lambda_L(x,y)$ and the paths of $\hat{\Lambda}_{L,m}(x,y) \in \mathcal{B}^I_\infty(\bar{I\!R}^2_+)$ can be (almost surely) decomposed into

$$\hat{\Lambda}^*_{L,m}\big(\frac{m}{k}G\big(G_m^{-1}\big(\frac{k}{m}x\big)\big), \frac{m}{k}H\big(H_m^{-1}\big(\frac{k}{m}x\big)\big)\big)$$

with $\frac{m}{k}G\big(G_m^{-1}\big(\frac{k}{m}x\big)\big)$ and $\frac{m}{k}H\big(H_m^{-1}\big(\frac{k}{m}x\big)\big)$ being the adjusted generalized inverse function (empirical quantile function of the tail) of $\hat{\Lambda}^*_{L,m}(x,\infty)$ and $\hat{\Lambda}^*_{L,m}(\infty,\, y)$. □

Corollary 4.5.4 (Covariance structure) *The covariance structure of* $G_{\Lambda^*_U}$ *and* $G_{\Lambda^*_L}$ *in Theorem* 4.5.2 *is given by*

$$I\!E\big(\mathbb{G}_{\hat{\Lambda}^*_U}(x,y)\cdot\mathbb{G}_{\hat{\Lambda}^*_U}(\bar{x},\bar{y})\big) = \Lambda_U\big(\min\{x,\bar{x}\},\min\{y,\bar{y}\}\big) \quad \text{and} \tag{4.26}$$

$$I\!E\big(\mathbb{G}_{\hat{\Lambda}^*_L}(x,y)\cdot\mathbb{G}_{\hat{\Lambda}^*_L}(\bar{x},\bar{y})\big) = \Lambda_L\big(\min\{x,\bar{x}\},\min\{y,\bar{y}\}\big) \tag{4.27}$$

for $(x,y)', (\bar{x},\bar{y})' \in \bar{I\!R}^2_+$.

Further the limiting process in Theorem 4.5.3 *can be expressed by*

$$\begin{aligned}\mathbb{G}_{\hat{\Lambda}_U}(x,y) &= \mathbb{G}_{\hat{\Lambda}^*_U}(x,y) \\ &- \frac{\partial}{\partial x}\Lambda_U(x,y)\mathbb{G}_{\hat{\Lambda}^*_U}(x,\infty) - \frac{\partial}{\partial y}\Lambda_U(x,y)\mathbb{G}_{\hat{\Lambda}^*_U}(\infty,y)\end{aligned} \tag{4.28}$$

and

$$\begin{aligned}\mathbb{G}_{\hat{\Lambda}_L}(x,y) &= \mathbb{G}_{\hat{\Lambda}^*_L}(x,y) \\ &- \frac{\partial}{\partial x}\Lambda_L(x,y)\mathbb{G}_{\hat{\Lambda}^*_L}(x,\infty) - \frac{\partial}{\partial y}\Lambda_L(x,y)\mathbb{G}_{\hat{\Lambda}^*_L}(\infty,y).\end{aligned} \tag{4.29}$$

Proof. The covariance structures (4.26) and (4.27) have been derived in the proof of Theorem 4.5.2. The second assertion follows from the Delta method (see Theorem 4.4.5). Notice that the derivative Φ'_{Λ_L} of the map Φ (at the point Λ_L) utilized in Theorem 4.5.3 is of the form

$$\Phi'_{\Lambda_L}(\alpha)(x,y) = \alpha(x,y) - \frac{\partial}{\partial x}\Lambda_L(x,y)\alpha(x,\infty) - \frac{\partial}{\partial y}\Lambda_L(x,y)\alpha(\infty,y)$$

according to the derivative map of the inverse operator (part ii) of the proof of Theorem 4.5.3, Lemma A (in the appendix), and the chain rule (Lemma 3.9.3 in Van der Vaart and Wellner (1996)). □

Corollary 4.5.5 (Asymptotic normality of $\hat{\lambda}_{U,m}$ and $\hat{\lambda}_{L,m}$) *Under the conditions of Theorem* 4.5.3

$$\sqrt{k}\big\{\hat{\lambda}_{U,m} - \lambda_U\big\} \xrightarrow{d} \mathcal{N}_{0,\sigma_U} \quad \text{and} \quad \sqrt{k}\big\{\hat{\lambda}_{L,m} - \lambda_L\big\} \xrightarrow{d} \mathcal{N}_{0,\sigma_L},$$

where $\mathcal{N}_{0,\sigma_U}$ *and* $\mathcal{N}_{0,\sigma_L}$ *are centered normal-distributed random variables with variances*

$$\begin{aligned}\sigma_U^2 &= \lambda_U + \Big(\frac{\partial}{\partial x}\Lambda_U(1,1)\Big)^2 + \Big(\frac{\partial}{\partial y}\Lambda_U(1,1)\Big)^2 \\ &+ 2\lambda_U\Big(\Big(\frac{\partial}{\partial x}\Lambda_U(1,1)-1\Big)\Big(\frac{\partial}{\partial y}\Lambda_U(1,1)-1\Big)-1\Big)\end{aligned} \tag{4.30}$$

and

$$\begin{aligned}\sigma_L^2 &= \lambda_L + \Big(\frac{\partial}{\partial x}\Lambda_L(1,1)\Big)^2 + \Big(\frac{\partial}{\partial y}\Lambda_L(1,1)\Big)^2 \\ &+ 2\lambda_L\Big(\Big(\frac{\partial}{\partial x}\Lambda_L(1,1)-1\Big)\Big(\frac{\partial}{\partial y}\Lambda_L(1,1)-1\Big)-1\Big).\end{aligned} \tag{4.31}$$

Proof. Note that e.g. for the lower tail-copula Λ_L we know from Corollary 4.5.4 that

$$\begin{aligned}\mathbb{E}\mathbb{G}^2_{\hat{\Lambda}_L}(x,y) &= \Lambda_L(x,y) + \Big(\frac{\partial}{\partial x}\Lambda_L(x,y)\Big)^2 x + \Big(\frac{\partial}{\partial y}\Lambda_L(x,y)\Big)^2 y \\ &+ 2\Lambda_L(x,y)\Big(\Big(\frac{\partial}{\partial x}\Lambda_L(x,y)-1\Big)\Big(\frac{\partial}{\partial y}\Lambda_L(x,y)-1\Big)-1\Big).\end{aligned}$$ □

Example. To illustrate the latter results we calculate the tail copula Λ_L and the asymptotic variance σ_L^2 in (4.31) for the well-known Pareto copula. See also Tables 3.2 and 3.3 for further copulae.

The bivariate Pareto copula $C(u,v)$ is given by

$$C(u,v) = \max\Big([u^{-\theta}+v^{-\theta}-1]^{-1/\theta},0\Big), \quad \theta\in[-1,\infty)\backslash\{0\}.$$

It has been shown (see Table 3.2) that the Pareto copula is lower tail-dependent with lower tail-dependence coefficient $\lambda_L = 2^{-1/\theta}$ for $\theta > 0$. Further, the lower tail copula exists for $\theta > 0$ and can be expressed by

$$\Lambda_L(x,y) = \big(x^{-\theta}+y^{-\theta}\big)^{-1/\theta}.$$

Thus, the partial derivatives are

$$\frac{\partial}{\partial x}\Lambda_L(x,y) = \big(x^{-\theta}+y^{-\theta}\big)^{-((1/\theta)+1)}x^{-(\theta+1)}$$

and

$$\frac{\partial}{\partial y}\Lambda_L(x,y) = \big(x^{-\theta}+y^{-\theta}\big)^{-((1/\theta)+1)}y^{-(\theta+1)}.$$

Consequently, the asymptotic variance σ_L^2 in (4.31) is given by (see also Figure 4.1)

$$\sigma_L^2 = 2^{-1/\theta} - \frac{3}{2}4^{-1/\theta} + \frac{1}{2}8^{-1/\theta}.$$

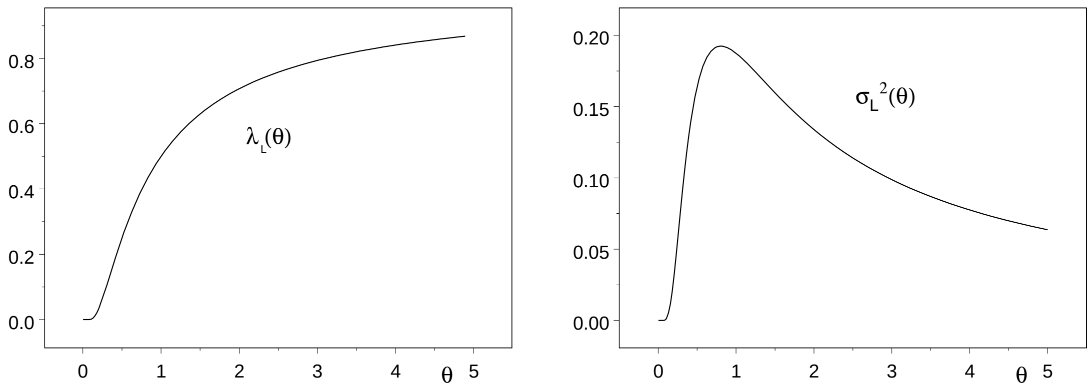


Figure 4.1: Lower tail-dependence coefficient $\lambda_L(\theta)$ (left plot) and corresponding asymptotic variance $\sigma_L^2(\theta)$ as in formula (4.31) (right plot) for the bivariate Pareto copula.

4.6 Strong consistency

Theorem 4.6.1 *Let F be a bivariate distribution function with continuous marginal distribution functions G and H. If the tail copulae $\Lambda_U \neq 0$ and $\Lambda_L \neq 0$ exist and $k/\log\log m \to \infty$ as $m \to \infty$ then $\hat{\Lambda}_{U,m}$ converges almost surely to Λ_U and $\hat{\Lambda}_{L,m}$ converges almost surely to Λ_L in the space $\mathbb{B}_\infty(\bar{I\!R}^2_+)$. In particular, for every $\varepsilon > 0$*

$$\mathbb{P}\big(\lim_{m\to\infty} d(\hat{\Lambda}_{U,m}, \Lambda_U) < \varepsilon\big) = 1 \quad \textit{and} \quad \mathbb{P}\big(\lim_{m\to\infty} d(\hat{\Lambda}_{L,m}, \Lambda_L) < \varepsilon\big) = 1. \tag{4.32}$$

Proof. We restrict the proof to the upper tail-copula. Recall that a sequence converges in the space $\mathbb{B}_\infty(\bar{I\!R}^2_+)$ with respect to the metric d (cf. (4.18)) if the sequence converges uniformly on each compact subset T_i introduced in Definition 4.4.2. Let $T > 0$ be an arbitrary but fixed constant. The conclusion now follows with the strong consistency result for empirical stable tail-dependence functions (defined in Section 2.1.2) given in Theorem 1.1 in Qi (1997) and the relationship $\Lambda_U(x,y) = x + y - l(x,y)$ (see also Section 2.2.4). Further, we utilize the fact that

$$|\hat{\Lambda}_{U,m}(x,y) - \Lambda_U(x,y)| = \Big|\frac{1}{k}\sum_{j=1}^{m} \mathbf{1}_{\{R_{m1}^{(j)} > m-kx \text{ and } R_{m2}^{(j)} > m-ky\}} - \Lambda_U(x,y)\Big| \leq$$

$$
\begin{aligned}
&\leq \Big|\frac{1}{k}\sum_{j=1}^{m}\mathbf{1}_{\{R_{m1}^{(j)}>m-kx \text{ or } R_{m2}^{(j)}>m-ky\}} - l(x,y)\Big| \\
&+ \Big|\frac{1}{k}\sum_{j=1}^{m}\mathbf{1}_{\{R_{m1}^{(j)}>m-kx\}} - x\Big| + \Big|\frac{1}{k}\sum_{j=1}^{m}\mathbf{1}_{\{R_{m1}^{(j)}>m-ky\}} - y\Big| \\
&\leq \Big|\frac{1}{k}\sum_{j=1}^{m}\mathbf{1}_{\{R_{m1}^{(j)}>m-kx \text{ or } R_{m2}^{(j)}>m-ky\}} - l(x,y)\Big| + \frac{2}{k}.
\end{aligned}
$$

The proof for the lower tail-copula is analogous. □

4.7 General rank order statistics for extreme events

In the present section we restrict ourself to the bivariate lower tail-copula Λ_L. Rank order statistics of the type

$$
\frac{1}{m}\sum_{j=1}^{m} J\big(R_{m1}^{(j)}/m, R_{m2}^{(j)}/m\big) = \frac{1}{m}\sum_{j=1}^{m} J(G_m(X^{(j)}), H_m(Y^{(j)}))
$$

have been investigated, for example, by Ruymgaart, Shorack, and van Zwet (1972), Ruymgaart (1974) and Rüschendorf (1976). Recently, Fermanian, Radulović, and Wegkamp (2002) considered general rank order statistics in the framework of empirical copula processes.

In the context of tail copulae a similar family of multivariate rank order statistics for extreme events can be investigated:

$$
\mathcal{R}_m := \frac{1}{k}\sum_{j=1}^{m} J\Big(R_{m1}^{(j)}/k, R_{m2}^{(j)}/k\Big). \tag{4.33}
$$

The next theorem establishes asymptotic normality of $\mathcal{R}_m$ under certain regularity assumptions on J.

Theorem 4.7.1 *Let F be a bivariate distribution function with continuous marginal distributions G and H. Further suppose that the (lower) tail copula $\Lambda_L \not\equiv 0$ exists and possesses continuous partial derivatives. Assume that $J : \bar{\mathbb{R}}_+^2 \to \mathbb{R}$ is of bounded variation, continuous from above with discontinuities of the first kind (Neuhaus 1971), and bounded on $\bar{\mathbb{R}}_+^2$. Then*

$$
\frac{1}{\sqrt{k}}\sum_{j=1}^{m}\Big(J\Big(\frac{R_{m1}^{(j)}}{k}, \frac{R_{m2}^{(j)}}{k}\Big) - \mathbb{E}J(G(X^{(j)}), H(Y^{(j)}))\Big) \xrightarrow{w} \int_{\bar{\mathbb{R}}_+^2} \mathbb{G}_{\Lambda_L}(x,y)dJ(x,y), \tag{4.34}
$$

where $\mathbb{G}_{\Lambda_L}$ is a centered continuous Gaussian field as in Theorem 4.5.3 and weak convergence takes place in $\mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)$. Moreover, the limiting process is also centered Gaussian.

Proof. Under the given prerequisites it follows that

$$\begin{aligned}
&\frac{1}{\sqrt{k}}\sum_{j=1}^{m}\Big(J\Big(\frac{R_{m1}^{(j)}}{k},\frac{R_{m2}^{(j)}}{k}\Big)-I\!\!E J(G(X^{(j)}),H(Y^{(j)}))\Big)\\
&=\frac{m}{\sqrt{k}}\int_{[0,1]^2}J\Big(\frac{m}{k}u,\frac{m}{k}v\Big)d(\bar{C}_m-C)(u,v)=:I_1
\end{aligned}$$

with $\bar{C}_m(u,v)=\frac{1}{m}\sum_{j=1}^{m}\mathbf{1}_{\{G_m(X^{(j)})\le u \text{ and } H_m(Y^{(j)})\le v\}}$ being the modified empirical copula process. Utilizing the integration by parts formula, stated in Baron, Liflyand, and Stadtmüller (2000), yields

$$\begin{aligned}
I_1 &= \int_{[0,1]^2}\frac{m}{\sqrt{k}}\Big((\bar{C}_m-C)(u-,v-)\Big)dJ\Big(\frac{m}{k}u,\frac{m}{k}v\Big)\\
&- \int_{[0,1]}\frac{m}{\sqrt{k}}((\bar{C}_m-C)(u-,1))dJ\Big(\frac{m}{k}u,1\Big)\\
&- \int_{[0,1]}\frac{m}{\sqrt{k}}((\bar{C}_m-C)(1,v-))dJ\Big(1,\frac{m}{k}v\Big)=:I_2.
\end{aligned}$$

Substituting $x=\frac{m}{k}v$ and $y=\frac{m}{k}u$ provides

$$\begin{aligned}
I_2 &= \int_{[0,m/k]^2}\frac{m}{\sqrt{k}}\Big((\bar{C}_m-C)\Big(\frac{k}{m}x-,\frac{k}{m}y-\Big)\Big)dJ(x,y)\\
&- \int_{[0,m/k]}\frac{m}{\sqrt{k}}\Big((\bar{C}_m-C)\Big(\frac{k}{m}x-,1\Big)\Big)dJ(x,1)\\
&- \int_{[0,m/k]}\frac{m}{\sqrt{k}}\Big((\bar{C}_m-C)\Big(1,\frac{k}{m}y-\Big)\Big)dJ(1,y)=:I_3.
\end{aligned}$$

Notice that

$$\begin{aligned}
\frac{m}{\sqrt{k}}\big((\bar{C}_m-C)(\frac{k}{m}x-,1)\big) &= \sqrt{k}\big(\hat{\Lambda}_{L,m}(x-,\infty)-x\big)\\
&= \sqrt{k}\big(\frac{1}{k}[kx-]-x\big)\in[0,1/\sqrt{k}].
\end{aligned}$$

Thus,

$$\frac{m}{\sqrt{k}}\big((\bar{C}_m-C)(\frac{k}{m}x-,1)\big)=O\Big(\frac{1}{\sqrt{k}}\Big).$$

The expression $\frac{m}{\sqrt{k}}\big((\bar{C}_m-C)(1,\frac{k}{m}y-)\big)$ possesses the same property. Therefore, the continuous mapping theorem (Van der Vaart and Wellner (1996), Theorem 1.3.6) leads to

$$I_3\xrightarrow{w}\int_{\bar{I\!\!R}_+^2}\mathbb{G}_{\Lambda_L}(x-,y-)dJ(x,y)=\int_{\bar{I\!\!R}_+^2}\mathbb{G}_{\Lambda_L}(x,y)dJ(x,y)$$

which is centered Gaussian. □

4.8 Multidimensional extensions

Again we consider only the lower (empirical) tail-copula Λ_L ($\hat{\Lambda}_L$). All results can be similarly stated for the upper (empirical) tail-copula.

Theorem 4.8.1 *Let F be a multivariate distribution function with continuous marginal distribution functions. Then the (lower) tail copula $\Lambda_L^{I,J}(x) \not\equiv 0$ exists if for some $j \in J$ the (lower) tail copula $\Lambda_L^{I\cup J\backslash\{j\},\{j\}}(x) \not\equiv 0$ exists. Moreover, (putting $0 := 0/0$)*

$$\Lambda_L^{I,J}(x) = \frac{\Lambda_L^{I\cup J\backslash\{j\},\{j\}}(x)}{\Lambda_L^{J\backslash\{j\},\{j\}}(x)}. \tag{4.35}$$

Proof. Suppose $\Lambda_L^{I\cup J\backslash\{j\},\{j\}}(x) \not\equiv 0$ exists for some $j \in J$. Then $\Lambda_L^{J\backslash\{j\},\{j\}}(x) > 0$ for all x according to the proof of part v) of Theorem 4.3.1. By Definition 2.2.7 we conclude that $\Lambda_L^{I,J}(x) \not\equiv 0$ exists and equation (4.35) holds. □

A nonparametric estimator for the lower tail copula is defined by

$$\hat{\Lambda}_{L,m}^{I,J}(x) := \sum_{j=1}^{m} \mathbf{1}_{\{R_{ml}^{(j)} \le kx_l, \forall l \in I \cup J\}} \Big/ \sum_{j=1}^{m} \mathbf{1}_{\{R_{ml}^{(j)} \le kx_l, \forall l \in J\}} \tag{4.36}$$

with $k = k(m) \to \infty$ and $k/m \to 0$ as $m \to \infty$.

Theorem 4.8.2 (Asymptotic normality) *Assume that the conditions of Theorem 4.8.1 are fulfilled, Λ_L possesses continuous partial derivatives, and the multivariate version of the Second order condition 4.5.1 holds. Then, for $\sqrt{k}A(m/k) \to 0$ as $m \to \infty$*

$$\sqrt{k}\{\hat{\Lambda}_{L,m}^{I,J}(x) - \Lambda_L^{I,J}(x)\} \xrightarrow{w} \mathbb{G}_{\Lambda_L}^{I,J}(x),$$

where $\mathbb{G}_{\Lambda_L}^{I,J}(x)$ is a centered tight continuous Gaussian random field, and weak convergence takes place in $\mathbb{B}_\infty(\bar{\mathbb{R}}_+^n\backslash[0,T)^n)$ for every $T > 0$.

The *proof* follows the same lines as the proof of Theorem 4.5.3.

Theorem 4.8.3 (Strong consistency) *Let F be a multivariate distribution function with continuous marginal distribution functions. Under the conditions of Theorem 4.8.1 and $k/\log\log m \to \infty$ as $m \to \infty$, $\hat{\Lambda}_{L,m}^{I,J}(x)$ converges almost surely to $\Lambda_L^{I,J}(x)$ in the space $\mathbb{B}_\infty(\bar{\mathbb{R}}_+^n\backslash[0,T)^n)$ for every $T > 0$.*

Proof. For some $j \in J$, put $I_1 := I \cup J \backslash \{j\}$ and $I_2 := J \backslash \{j\}$. Let $k/\log\log m \to \infty$. Then

$$\begin{aligned}
&|\hat{\Lambda}_{L,m}^{I,J}(x) - \Lambda_L^{I,J}(x)| \\
&\quad = |\hat{\Lambda}_{L,m}^{I_1,\{j\}}(x) - \Lambda_L^{I_1,\{j\}}(x)|/\Lambda_L^{I_2,\{j\}}(x) \\
&\qquad + \frac{\hat{\Lambda}_{L,m}^{I_1,\{j\}}(x)}{\hat{\Lambda}_{L,m}^{I_2,\{j\}}(x)} \frac{1}{\Lambda_L^{I_2,\{j\}}(x)} \left| \hat{\Lambda}_{L,m}^{I_2,\{j\}}(x) - \Lambda_L^{I_2,\{j\}}(x) \right| \\
&\quad =: C_{1,m}(x) + C_{2,m}(x).
\end{aligned}$$

Regarding the first term: $C_{1,m}(x) \to 0$ as $m \to \infty$ almost surely. This follows essentially from Theorem 1.2 in Qi (1997) and the fact that for every $T > 0$, the estimator $\hat{\Lambda}_{L,m}$ is bounded from above and below by strictly positive constants on $T_i \backslash [0, T)^n$ (cf. Definition 4.4.2). The second term: $C_{2,m}(x) \to 0$ as $m \to \infty$ almost surely for the same reasons. □

In Section 4.9.3, a simulation study compares the sample behavior of the latter estimators.

4.9 Estimation of tail dependence for elliptically contoured distributions

The present section and the next one primarily serve as a summary of various estimators for the tail-dependence coefficient. Statistical methods testing for tail dependence or tail independence go beyond the scope of this work, for that we refer the reader to Ledford and Tawn (1996) and Draisma, Drees, Ferreira, and de Haan (2001). For the remaining of this chapter, the term tail-dependence coefficient is abbreviated as TDC. For further details and results we refer to Frahm, Junker, and Szimayer (2003).

According to Proposition 2.2.5, the TDCs λ_L and λ_U are limiting copula properties. Thus we may write $\lambda_L = \lambda_L(C)$ ($\lambda_U = \lambda_U(C)$) or $\lambda_L = \lambda_L(\theta)$ ($\lambda_U = \lambda_U(\theta)$) if C is a member of a parametric copula-family C_θ with parameter vector θ. For the sake of simplicity we write λ for the TDC whenever $\lambda_L = \lambda_U$.

The main difficulties in estimating the TDC arise in the limited availability of extreme data. Therefore we suggest several parametric and nonparametric estimators ascribed either to the entire data set or to a subset of extreme data. Regarding the latter, EVT is the natural choice for inferences on extreme events of random vectors or the tail behavior of probability distributions. According to Sections 2.1.2 and 3.3.2, we can approximate the tail of a probability distribution by an appropriate extreme value distribution. In the one-dimensional setting the class of extreme value distributions has a solely parametric representation (see Theorem 2.1.1), so it suffices to apply parametric estimation methods. By contrast, multidimensional extreme value distributions are characterized by a parametric and a nonparametric component (cf. Section 3.3.2). This leads to more complicated estimation methods. Parametric estimation methods have the advantage of being efficient given that the model is true whereas nonparametric estimation methods avoid model-misclassifications.

4.9.1 Estimation from the entire sample

In the present section we assume that the observed data arise from an elliptically contoured random sample $X, X^{(1)}, \ldots, X^{(m)},\ m \in \mathbb{N}$. If not stated otherwise, suppose $X \in E_n(\mu, \Sigma, \Phi)$ with positive-definite dispersion matrix Σ. Then a nonsingular matrix $\sqrt{\Sigma}$ exists, i.e., the Cholesky root of Σ with inverse $\sqrt{\Sigma}^{-1}$. According to equation (3.16), the following stochastic representation exists

$$X \stackrel{d}{=} \mu + R_n \sqrt{\Sigma}' U^{(n)}. \tag{4.37}$$

In Section 3.2.2 we essentially showed that the coinciding upper and lower TDC of two components of an elliptical random vector X, say X_i and X_j, depends only on two parameters in the case of tail dependence. Namely, the TDC depends on the tail index α of the corresponding density generator or generating distribution function, if they are regularly varying (see formula (3.18)), and on the correlation coefficient ρ_{ij} between X_i and X_j. In particular, $\lambda_{ij} = \lambda_{ij}(\alpha, \rho_{ij})$. Observe that the correlation coefficient may not exist; however the expression $\rho_{ij} = \Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}$ is still well defined. In this case we refer to ρ_{ij} as some dependence parameter. If not stated otherwise, ρ_{ij} is always referred to as the correlation coefficient for notational convenience.

For elliptical random vectors X, regular variation of the generating variate R_n is equivalent to regular variation of X according to Section 2.3.2 and the proof of Theorem 3.3.2. Moreover, the tail index α of the generating variate R_n equals the tail index of the elliptical random vector X. Thus, in the following we may always refer to the tail index α of X.

A closed form expression for the TDC of elliptical distributions is given via formula (3.18). Notice that this formula can be rewritten as

$$\lambda_{ij} = 2 \cdot \bar{t}_{1-\alpha}\left(\sqrt{1-\alpha} \cdot \sqrt{\frac{1-\rho_{ij}}{1+\rho_{ij}}}\right), \qquad -\alpha > 0, \tag{4.38}$$

where $\bar{t}_{1-\alpha}$ denotes the survival function of the Student's univariate t-distribution with $1-\alpha$ degrees of freedom (cf. Frahm, Junker, and Szimayer (2003)). Thus, the upper and lower TDC between X_i and X_j can be estimated via formula (4.38) if some estimates of the tail index α and the correlation coefficient ρ_{ij} are given. In particular we suggest

$$\hat{\lambda}^E_{ij} = \hat{\lambda}^E_{ij,m} = \lambda_{ij}(\hat{\alpha}_m, \hat{\rho}_{ij,m}). \tag{4.39}$$

In the following, a method is presented which extracts approximate realizations of the generating variate R_n, and provides a powerful estimator for the correlation matrix $\rho = (\rho ij)_{i,j=1,\ldots,n}$. Moreover, once approximate realizations of R_n are given, we can fit an appropriate distribution for R_n.

Due to the stochastic representation (4.37) the following inverse relation holds

$$R_n U^{(n)} \stackrel{d}{=} \sqrt{\Sigma}'^{-1} (X - \mu) .$$

Hence,

$$\delta O R_n U^{(n)} \stackrel{d}{=} \delta O \sqrt{\Sigma}'^{-1} (X - \mu) = \Gamma (X - \mu) , \quad \text{with } \delta > 0, \ \Gamma := \delta O \sqrt{\Sigma}'^{-1}, \qquad (4.40)$$

where δ denotes an arbitrary scaling constant and O refers to any orthogonal matrix.

Given some location vector μ and some inversion matrix Γ, we may sieve realizations from the spherical random vector $\Gamma (X - \mu)$ and from the rescaled generating variate δR_n since

$$\|\Gamma (X - \mu)\|_2 \stackrel{d}{=} \left\|\delta O R_n U^{(n)}\right\|_2 = \delta R_n \cdot \left\|O U^{(n)}\right\|_2 = \delta R_n =: R_n^*. \qquad (4.41)$$

Lindskog (2000) shows that the heavier the tails of the one-dimensional margins of X the less is Pearson's standard covariance estimator suitable for estimating the scaling matrix Σ or the correlation matrix ρ. The covariances of the margins of X may even be infinite if the tail index $-\alpha$ of X is smaller than 2. In this case, Pearson's covariance estimator is vulnerable to contaminated and outlier data. Moreover, even the sample mean vector is not an appropriate estimator for μ if $-\alpha < 1$ because, in that case, the expectation does not exist. Hence, robust estimation methods for both the location parameter μ and the correlation matrix ρ are required. This task is tackled in the remaining section.

Since elliptical distributions are symmetric, the median vector $\mu_{0.5}$ corresponds to the location vector μ. Under some technical condition given in Büning and Trenkler (1978), p. 61, the sample-median $\overline{x}_{0.5}$ is a consistent estimator for μ. So $\widehat{\mu} = \widehat{\mu}_m = \overline{x}_{0.5}$ seems to be an appropriate estimator for μ (we drop the subscript m if it is clear from the context). Furthermore, let $\widehat{\Gamma}$ be any estimator of the inversion matrix Γ. Due to (4.40) and (4.41) the relation

$$O U^{(n)} \stackrel{d}{=} \frac{\Gamma (X - \mu)}{\|\Gamma (X - \mu)\|_2} \approx \frac{\widehat{\Gamma} (X - \widehat{\mu})}{\left\|\widehat{\Gamma} (X - \widehat{\mu})\right\|_2}$$

holds if $\widehat{\Gamma}$ is a suited estimator for Γ and the sample size m is sufficiently large. The random vector $U^{(n)}$ satisfies the rotation invariance property, i.e., for every orthogonal matrix O the random vector $OU^{(n)}$ equals in distribution to $U^{(n)}$ which is uniformly distributed on the unit sphere.

Let

$$u_i^* := \frac{\Gamma^* (x_i - \widehat{\mu})}{\|\Gamma^* (x_i - \widehat{\mu})\|_2}, \qquad i = 1, ..., m,$$

where Γ^* is any nonsingular matrix and $x_1, ..., x_m$ are realizations of X. The main goal is to find a matrix Γ^* which leads to the most "uniformly" distributed vectors $u_1^*, ..., u_m^*$ on the unit sphere. "Uniformity" means in this context that there are no areas on the unit sphere with significantly high vector concentration on the one hand and no areas with significantly low vector concentration on the other hand. In order to measure vector concentrations we have to constitute an appropriate distance measure $D_{kj}(\Gamma^*)$ for any given vectors u_k^* and u_j^* like e.g. the Euclidean metric

$$\left\| u_k^* - u_j^* \right\|_2 \tag{4.42}$$

or the angle between the vectors, i.e.,

$$\arccos \left\langle u_k^*, u_j^* \right\rangle , \tag{4.43}$$

where $\langle .,. \rangle$ denotes the scalar-product. The distance between the vector u_k^* and the nearest vector next to it is defined by

$$D_k(\Gamma^*) := \min_{\{j:\ j \neq k\}} D_{kj}(\Gamma^*) .$$

For a better understanding we concentrate on the bivariate case $n = 2$, and take the angle between the vectors as a distance measure. The generalization to the multivariate case is discussed in Frahm and Junker (2003). In the bivariate context, perfect uniformity is simply characterized by the unit circle divided into m equidistant intervals. In particular, every distance $D_i(\Gamma^*) = \min_{\{j:\ j \neq i\}} \arccos \left\langle u_i^*, u_j^* \right\rangle,\ i = 1, ..., m,$ should approximately correspond to $2\pi/m$. Consequently, an estimator for Γ may be specified by

$$\widehat{\Gamma} := \widehat{\Gamma}_m = \arg\min_{\Gamma^*} \sum_{i=1}^{m} \left(D_i(\Gamma^*) - \frac{2\pi}{m} \right)^2 .$$

Notice that the distances between all vectors remain unchanged under linear transformations $\delta O u^*$. Therefore, $\widehat{\Gamma}$ is uniquely defined up to rotations and scale transformations. For reasons of uniqueness we constrain the matrix Γ^* by the following two conditions.

(i) Γ^* is a lower triangular matrix, and

(ii) the last column vector of Γ^* is a unit vector.

With $\Gamma := \delta O \sqrt{\Sigma}'^{-1}$ and $O^{-1} = O'$ we obtain

$$\begin{aligned} \Gamma^{-1} \left(\Gamma^{-1} \right)' &= \left(\delta O \sqrt{\Sigma}'^{-1} \right)^{-1} \left(\left(\delta O \sqrt{\Sigma}'^{-1} \right)^{-1} \right)' \\ &= \frac{1}{\delta^2} \cdot \left(\sqrt{\Sigma}' O' \right) \left(O \sqrt{\Sigma} \right) = \frac{1}{\delta^2} \cdot \sqrt{\Sigma}' \sqrt{\Sigma} = \frac{\Sigma}{\delta^2}. \end{aligned} \tag{4.44}$$

Since $\widehat{\Gamma} \approx \Gamma$, the rescaled dispersion matrix $\Sigma/\delta^2 =: \Sigma^*$ can be estimated by

$$\widehat{\Gamma}^{-1} \left(\widehat{\Gamma}^{-1} \right)' =: \widehat{\Sigma}^*.$$

Note that $\widehat{\Gamma}$ is a triangular matrix and so the inverse of $\widehat{\Gamma}$ exists if and only if the product of the diagonal elements of $\widehat{\Gamma}$ do not vanish, i.e., diag$(\widehat{\Gamma})$ has no element which equals zero. Hence, another constraint for Γ^* is that

(iii) the diagonal of Γ^* contains no element which equals zero.

The latter constraint ensures that $\widehat{\Sigma}^*$ is positive-definite since both, the matrix $\widehat{\Gamma}^{-1}$ and therewith its transpose, are nonsingular.

The transformation matrix $\rho = (\rho_{ij})$ is defined by means of reparameterization,

$$(\rho_{ij}) = \left(\frac{\Sigma_{ij}}{\Sigma_i \Sigma_j} \right),$$

with $\Sigma_i := \sqrt{\Sigma_{ii}}$, $i = 1, ..., n$, where $(\Sigma_{ij}) = \Sigma$. Because the scaling constant δ does not affect the matrix ρ, it can be estimated by

$$\widehat{\rho} = (\widehat{\rho}_{ij}) = \left(\frac{\widehat{\Sigma}^*_{ij}}{\widehat{\Sigma}^*_i \widehat{\Sigma}^*_j} \right), \tag{4.45}$$

with $\widehat{\Sigma}^*_i := \sqrt{\widehat{\Sigma}^*_{ii}}$, $i = 1, ..., n$, and $\left(\widehat{\Sigma}^*_{ij}\right) = \widehat{\Sigma}^*$.

Various estimation methods for the correlation matrix ρ of elliptical random vectors are available in the literature; for an overview see Lindskog (2000). The minimum volume estimator, for example, minimizes the volume of the ellipsoid generated by the Mahalanobis distance, i.e.,

$$\left\{ x \in I\!R^d : \sqrt{(x - \widehat{\mu})' \, \widehat{\Sigma}^{-1} \, (x - \widehat{\mu})} \leq \mathcal{M} \right\},$$

covering a certain number of data points $k \leq m$. The difficulty is to find the "right" k. If the number of covered data points is too large or too small, extreme values lead to high variance of the estimator. In other words, estimation methods which are based on the Mahalanobis distance are very sensitive to outliers. However, separating the correlation structure and the generating variate of elliptically distributed random vectors (like presented before) enables an elimination of extremes without reducing the data set and yields a robust correlation estimator. Moreover, the latter correlation estimator is always positive definite; which is an important advantage as erroneous transformations can be avoided.

After treating the estimation of the correlation coefficient, we now turn to the estimation of the tail index α. Due to equation (4.41), approximate realizations of the rescaled generating variate $R^*_n = \delta R_n$ are given by

$$\widehat{r}^*_i := \left\| \widehat{\Gamma} \, (x_i - \widehat{\mu}) \right\|_2, \qquad i = 1, ..., m.$$

However, the random variable R^*_n has not necessarily the same distribution as R_n, but the tail index of R_n is invariant under scale transformations. Hence, for estimations of the tail index α from the entire data sample it suffices to fit an appropriate distribution function F_θ to the realizations $\widehat{r}^*_i$, $i = 1, ..., m$. Finally observe that the tail index α is usually a function of θ which is determined by the distribution F_θ. Therefore, an estimate of α can be derived via some functional $\alpha(\widehat{\theta})$.

4.9.2 Estimation from the distribution's tail or extreme samples

Alternatively, the tail index α is now estimated solely from extreme realizations. Assume that the elliptical random vector X is regularly varying with index α. Then the corresponding TDCs are given via formula (4.38) and depend only on the tail index α and the correlation matrix ρ. The estimation of the tail index α reduces to a one-dimensional problem because the random variable $||X||_2$ is also regularly varying, and possesses the same tail index as X according to the proof of Theorem 3.3.2. Hence, we estimate the tail index with standard methods from univariate EVT utilizing the extreme values of $||X||_2$.

In Section 4.9.1, we introduced a method for separating the correlation structure and the random variate $\mathcal{R}_n \stackrel{d}{=} ||X||_2$ of an elliptically distributed random vector; precisely we obtained a rescaled random variate R_n^*. Thus, alternatively the tail index α can be estimated from extreme values of R_n^* instead from $\mathcal{R}_n \stackrel{d}{=} ||X||_2$ as the rescaling has non influence on the tail index α. However, it might have some influence regarding the finite sample estimation (see Section 4.11.1 for an graphical illustrations of the latter).

Different methods for estimating the (univariate) tail index are discussed in Embrechts, Klüppelberg, and Mikosch (1997). In contrast to the peaks over threshold and the block maxima approach, estimation methods based on upper order statistics (e.g. the Hill-estimator) do not require an estimate for the scale parameter of the generalized Pareto distribution or the generalized extreme value distribution, respectively. The Hill estimator represents a quite natural estimator for the tail index α. Namely,

$$\hat{\alpha}_m = \Big(\frac{1}{k}\sum_{j=1}^{k} \log ||X_{(j,m)}||_2 - \log ||X_{(k,m)}||_2\Big)^{-1}, \tag{4.46}$$

where $||X_{(j,m)}||_2$ denotes the j-th order statistics of $||X^{(1)}||_2, \ldots, ||X^{(m)}||_2$ and $k = k(m) \to \infty$ is appropriately chosen; for a discussion on the right choice we refer the reader to Embrechts, Klüppelberg, and Mikosch (1997), pp. 341. For our purposes, a Hill-type estimator with optimal sample fraction proposed by Drees and Kaufmann (1998) turns out to be satisfying.

The quantlet **TailCoeffEstimElliptical** in the mathematical software-program **Xplore** estimates the coinciding upper and lower TDCs of a bivariate elliptically distributed data-set via the estimator $\hat{\lambda}_m^E$ (see formula (4.39)) utilizing the Hill estimator (4.46). The inputs of this quantlet are the two-dimensional data-set and the threshold k. The result of the application is assigned to the variable `TDC` and contains the estimate for the coinciding lower and upper TDC.

```
TDC = TailCoeffEstimElliptical (data,threshold)
```

Quantlet 3.1: Estimates the upper and lower tail-dependence coefficient via formulae (4.39) and (4.46).

Further, we consider the circumstance that the random sample $X^{(1)}, \dots, X^{(m)}$ is not elliptically distributed but possesses an elliptical copula C. Assume that the distribution of the sample has continuous marginal distribution functions. Recall that in the presence of tail dependence Theorem 3.2.5 and Theorem 3.2.2 justify to consider only elliptical copulae with regularly varying generating variate or regularly varying density generator. If X is in the domain of attraction of some extreme value distribution, we may estimate the TDC using the homogeneity property (2.26) and the spectral measure representation (3.30) arising from the limiting extreme value distribution. Einmahl, de Haan, and Piterbarg (2001) and Einmahl, de Haan, and Sinha (1997) propose a nonparametric and a semi-parametric estimator for the spectral measure of an extreme value distribution. If the hypothesis that X is in the domain of attraction of an extreme value distribution is rejected, the TDC could still be estimated via the empirical tail-copula introduced in Section 4.2. Precisely, with C_m being the empirical copula

$$C_m(u,v) = F_m(F_{m1}^{-1}(u), F_{m2}^{-1}(v)) \tag{4.47}$$

and $R_{m1}^{(j)}$ and $R_{m2}^{(j)}$ being the rank of $X_1^{(j)}$ and $X_2^{(j)}$, $j = 1, \dots, m$, respectively, the first set of estimators for the upper and lower TDC are

$$\hat{\lambda}_{U,m} = \frac{m}{k} C_m\Big(\Big(1 - \frac{k}{m}, 1\Big] \times \Big(1 - \frac{k}{m}, 1\Big]\Big) = \frac{1}{k}\sum_{j=1}^{m} \mathbf{1}_{\{R_{m1}^{(j)} > m-k \text{ and } R_{m2}^{(j)} > m-k\}} \tag{4.48}$$

and

$$\hat{\lambda}_{L,m} = \frac{m}{k} C_m\Big(\frac{k}{m}, \frac{k}{m}\Big) = \frac{1}{k}\sum_{j=1}^{m} \mathbf{1}_{\{R_{m1}^{(j)} \le k \text{ and } R_{m2}^{(j)} \le k\}}, \tag{4.49}$$

where $k = k(m) \to \infty$ and $k/m \to 0$ as $m \to \infty$. The first expression in (4.48) has to be understood as the empirical copula-measure of the interval $(1 - k/m, 1] \times (1 - k/m, 1]$ (cf. Section 4.2). The second type of estimator is

$$\begin{aligned} \hat{\lambda}_{U,m}^{EVT} &= 2 - \frac{m}{k}\Big(1 - C_m\Big(1 - \frac{k}{m}, 1 - \frac{k}{m}\Big)\Big) \\ &= 2 - \frac{1}{k}\sum_{j=1}^{m} \mathbf{1}_{\{R_{m1}^{(j)} > m-k \text{ or } R_{m2}^{(j)} > m-k\}}, \end{aligned} \tag{4.50}$$

with $k = k(m) \to \infty$ and $k/m \to 0$ as $m \to \infty$. $\hat{\lambda}_{L,m}^{EVT}$ is similarly defined. The optimal choice of k is related to the usual variance-bias problem.

Asymptotic normality and strong consistency are addressed in Section 4.5 and Section 4.6. However, the next theorem provides another approach to prove strong consistency of the estimators $\hat{\lambda}_{U,m}$ and $\hat{\lambda}_{L,m}$ within the context of elliptical copulae. Asymptotic normality can be shown similarly.

Theorem 4.9.1 *Let X be a bivariate random vector with elliptical copula C having a regularly varying density generator or regularly varying generating variate. Let $\hat{\lambda}_{U,m}$ and $\hat{\lambda}_{L,m}$ be the upper and lower tail-dependence coefficient estimator given in (4.48) and (4.49). If $k = k(m) \to \infty$, $k/m \to 0$, $k/\log(\log m) \to \infty$ as $m \to \infty$ then*

$$\hat{\lambda}_m \to \lambda \quad \textit{almost surely} \quad \textit{as} \quad m \to \infty.$$

Proof. According to Stute (1984), p. 371, the distribution of C_m in (4.47) does not depend on the marginal distributions F_1 and F_2 so that w.l.o.g we may assume that $F_i,\ i = 1, 2$, are uniform distributions on the unit interval. Therefore, we are in the copula framework. Theorem 3.3.5 yields that C is in the domain of attraction of an extreme value distribution. The strong consistence is now a special case of Theorem 1.1 in Qi (1997) because of the uniform convergence of $C^m(1 - 1/(mx_1), \ldots, 1 - 1/(mx_n))$ to its corresponding extreme value distribution. □

The quantlet **TailCoeffEstimation** in the mathematical software-program **Xplore** estimates the upper and lower TDC based on formulae (4.48) and (4.49). The inputs of this quantlet are the two-dimensional data-set, the threshold k, and the instruction whether the lower or upper TDC is calculated. The result of the application is assigned to the variable `TDC` and contains the estimated lower or upper TDC, respectively.

```
TDC = TailCoeffEstimation(data,threshold,upper)
```

Quantlet 3.2: Estimates the upper and lower tail-dependence coefficient based on formulae (4.48) and (4.49).

4.9.3 Some empirical results

In this section we investigate some properties of the above introduced TDC estimators. 1000 independent copies of $m = 500, 1000$, 2000 iid random vectors of a bivariate standard t-distribution with $\theta = 1.5, 2, 3$ degrees of freedom are generated and the upper TDC's are estimated. The empirical bias and mean-squared error (MSE) for three different TDC estimation methods are derived and presented in Table 4.1, 4.2, 4.3, 4.4, and 4.5, respectively. Regarding the parametric estimation, we apply the procedure explained in Section 4.9.2, and estimate ρ by a trimmed Pearson-correlation coefficient with trimming portion 0.05% and θ by a Hill-type estimator. Here we choose the optimal threshold value k according to Drees and Kaufmann (1998). Observe that the common "correlation" coefficient ρ does not exist for $\theta \leq 2$. In this case, ρ denotes some dependence parameter and a more robust estimation procedure is developed in Section 4.9.1. However, the simulation study reveals that for t-distributions the behavior of the Pearson-correlation coefficient is quite satisfying, also for $\theta \leq 2$.

Original parameters	$\theta = 1.5$ $\lambda_U = 0.2296$	$\theta = 2$ $\lambda_U = 0.1817$	$\theta = 3$ $\lambda_U = 0.1161$
Estimator	$\hat{\lambda}_U$ Bias (MSE)	$\hat{\lambda}_U$ Bias (MSE)	$\hat{\lambda}_U$ Bias (MSE)
$m = 500$	0.0255 (0.00369)	0.0434 (0.00530)	0.0718 (0.00858)
$m = 1000$	0.0151 (0.00223)	0.0287 (0.00306)	0.0518 (0.00466)
$m = 2000$	0.0082 (0.00149)	0.0191 (0.00169)	0.0369 (0.00270)

Table 4.1: Empirical bias and MSE for the nonparametric upper TDC estimator $\hat{\lambda}_U$ given in (4.48).

Original parameters	$\theta = 1.5$ $\lambda_U = 0.2296$	$\theta = 2$ $\lambda_U = 0.1817$	$\theta = 3$ $\lambda_U = 0.1161$
Estimator	$\hat{\lambda}_U^{EVT}$ Bias (MSE)	$\hat{\lambda}_U^{EVT}$ Bias (MSE)	$\hat{\lambda}_U^{EVT}$ Bias (MSE)
$m = 500$	0.0539 (0.00564)	0.0703 (0.00777)	0.1031 (0.01354)
$m = 1000$	0.0333 (0.00301)	0.0491 (0.00437)	0.0748 (0.00744)
$m = 2000$	0.0224 (0.00173)	0.0329 (0.00228)	0.0569 (0.00436)

Table 4.2: Empirical bias and MSE for the nonparametric upper TDC estimator $\hat{\lambda}_U^{EVT}$ given in (4.50).

Original parameters	$\theta = 1.5$ $\lambda_U = 0.2296$	$\theta = 1.5$ $\alpha = 1.5$	$\theta = 1.5$ $\rho = 0$
Estimator	$\hat{\lambda}_U^E$ Bias (MSE)	$\hat{\alpha}$ Bias (MSE)	$\hat{\rho}$ Bias (MSE)
$m = 500$	0.0016 (0.00093)	0.00224 (0.04492)	0.00002 (0.00379)
$m = 1000$	0.0024 (0.00050)	−0.0147 (0.02354)	−0.0003 (0.00202)
$m = 2000$	0.0024 (0.00024)	−0.0157 (0.01023)	0.00075 (0.00103)

Table 4.3: Empirical bias and MSE for the parametric upper TDC estimator $\hat{\lambda}_U^E$ given in (4.39), the tail index estimator $\hat{\alpha}$, and the correlation coefficient estimator $\hat{\rho}$.

Original parameters	$\theta = 2$, $\lambda_U = 0.1817$	$\theta = 2$, $\alpha = 2$	$\theta = 2$, $\rho = 0$
Estimator	$\hat{\lambda}_U^E$	$\hat{\alpha}$	$\hat{\rho}$
	Bias (MSE)	Bias (MSE)	Bias (MSE)
$m = 500$	0.0035 (0.00095)	−0.0198 (0.10417)	−0.00261 (0.00339)
$m = 1000$	0.0058 (0.00057)	−0.0485 (0.05552)	0.0010 (0.00177)
$m = 2000$	0.0054 (0.00029)	−0.0606 (0.02994)	−0.00126 (0.00086)

Table 4.4: Empirical bias and MSE for the parametric upper TDC estimator $\hat{\lambda}_U^E$ given in (4.39), the tail index estimator $\hat{\alpha}$, and the correlation coefficient estimator $\hat{\rho}$.

Original parameters	$\theta = 3$, $\lambda_U = 0.1161$	$\theta = 3$, $\alpha = 3$	$\theta = 3$, $\rho = 0$
Estimator	$\hat{\lambda}_U^E$	$\hat{\alpha}$	$\hat{\rho}$
	Bias (MSE)	Bias (MSE)	Bias (MSE)
$m = 500$	0.0162 (0.00115)	−0.2219 (0.29568)	0.0021 (0.00319)
$m = 1000$	0.0154 (0.00076)	−0.2422 (0.20047)	0.0006 (0.00156)
$m = 2000$	0.0124 (0.00046)	−0.2175 (0.12921)	−0.00175 (0.00074)

Table 4.5: Empirical bias and MSE for the parametric upper TDC estimator $\hat{\lambda}_U^E$ given in (4.39), the tail index estimator $\hat{\alpha}$, and the correlation coefficient estimator $\hat{\rho}$.

Finally, we provide some plots illustrating the (non-) parametric estimation results for the upper TDC λ_U. Each plot presents 3×1000 TDC estimations with $m = 500, 1000, 2000$ in a consecutive ordering for $\theta = 2$. The plots visualize the decreasing empirical variance for increasing m.

The empirical study shows that the TDC estimator $\hat{\lambda}_U^E$ outperforms the other estimators. For $m = 2000$ the bias (MSE) of $\hat{\lambda}_U$ is three (five) times larger than the bias (MSE) of $\hat{\lambda}_U^E$ whereas the bias (MSE) of $\hat{\lambda}_U^{EVT}$ is two (one and a half) times larger than the bias (MSE) of $\hat{\lambda}_U$. However, remember that the estimator $\hat{\lambda}_U^E$ has been developed especially for elliptically contoured distributions. Thus, the estimator $\hat{\lambda}_U$ is recommended for TDC estimations of general or unknown bivariate distributions.

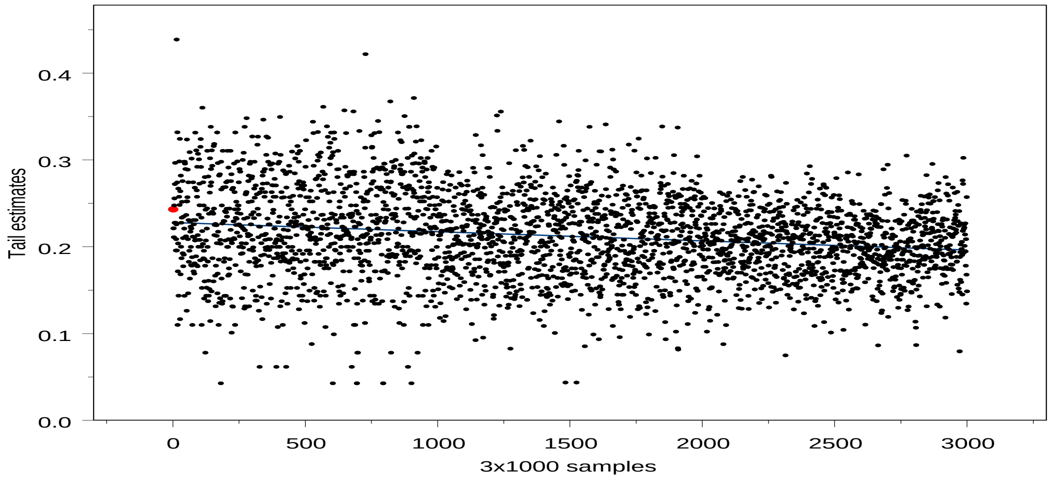


Figure 4.2: Nonparametric upper TDC estimates $\hat{\lambda}_U$ (see formula (4.48)) for 3×1000 iid samples of size $m = 500, 1000, 2000$ from a bivariate t-distribution with parameters $\theta = 2$, $\rho = 0$, and $\lambda_U = 0.1817$.

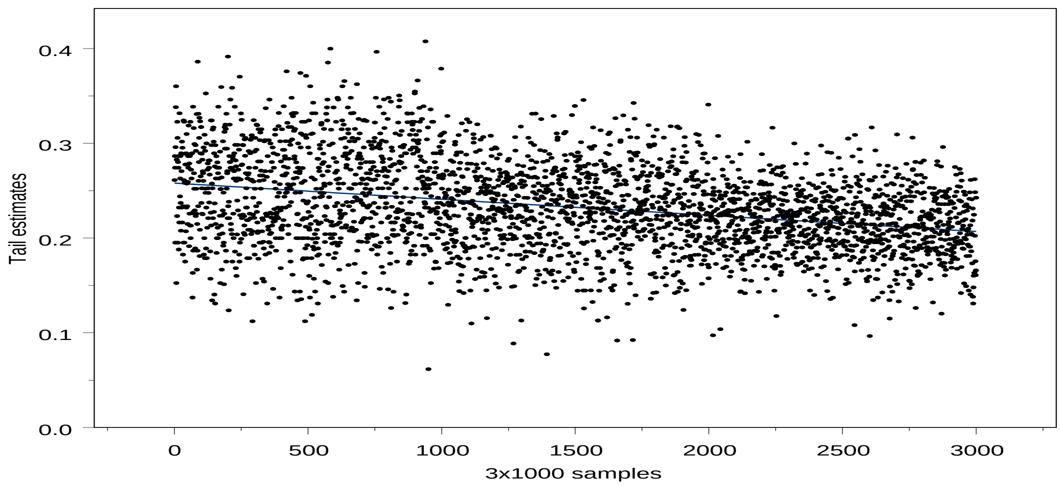


Figure 4.3: Nonparametric upper TDC estimates $\hat{\lambda}_U^{EVT}$ (see formula (4.50)) for 3×1000 iid samples of size $m = 500, 1000, 2000$ from a bivariate t-distribution with parameters $\theta = 2$, $\rho = 0$, and $\lambda_U = 0.1817$.

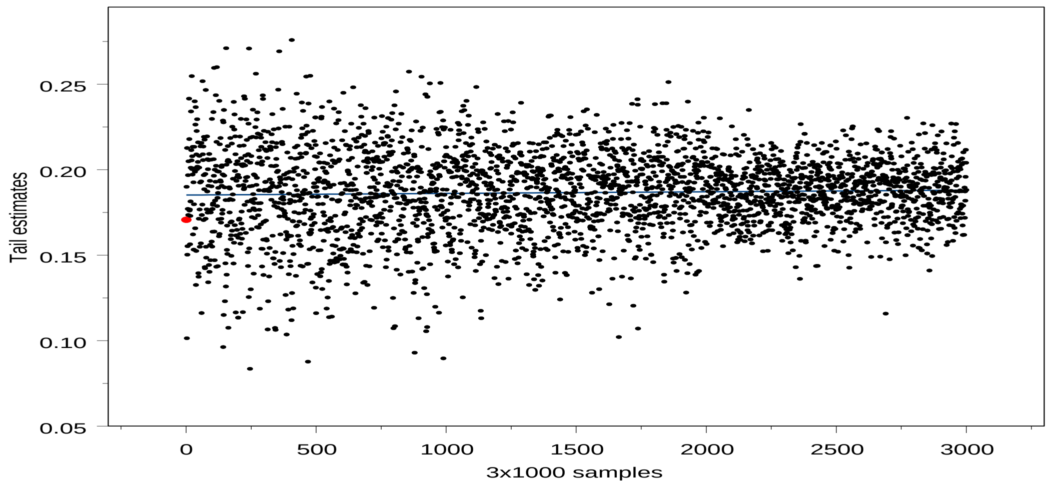


Figure 4.4: Parametric upper TDC estimates $\hat{\lambda}_U^E$ (see formula (4.39)) for 3×1000 iid samples of size $m = 500, 1000, 2000$ from a bivariate t-distribution with parameters $\theta = 2$, $\rho = 0$, and $\lambda_U = 0.1817$.

4.10 Tail dependence: Parametric versus nonparametric estimation

4.10.1 Estimation from the entire sample

In this section the lower and upper TDC are estimated from the entire sample or, in other words, from the entire distribution.

(Semi-) Parametric methods

If not stated otherwise, for the remaining part of the present chapter we concentrate only on the upper TDC λ_U and for convenience we drop the subscript U, i.e., $\lambda = \lambda_U$. One approach to model tail dependence is to choose a *parametric* copula family C_θ allowing for tail dependence and being appropriate to describe the data. Then, the TDC λ is estimated by some functional estimator $\widehat{\lambda} := \widehat{\lambda}_m = \lambda(\widehat{\theta}_m)$, $m \in I\!N$ (we drop the subscript m if it is clear from the context). Starting from a random sample $X^{(1)}, \ldots, X^{(m)}$ with common n-variate distribution function F, we may first estimate the marginal distribution functions $F_1, \ldots, F_n$ in order to obtain approximate realizations of the copula C_θ via

$$\widehat{U}^{(i)} = \left(\widehat{U}_1^{(i)}, ..., \widehat{U}_n^{(i)}\right)' = \left(\widehat{F}_1\left(X_1^{(i)}\right), ..., \widehat{F}_n\left(X_n^{(i)}\right)\right)', \; i = 1, \ldots, m.$$

Parametric and nonparametric procedures are distinguished to estimate the marginal distribution functions $F_1, \ldots, F_n$. Regarding the parametric procedure, we assume that n parameter vectors $p_1, ..., p_n$ determine the marginal distribution functions $F_1, \ldots, F_n$. In contrast, within the nonparametric procedure, all marginal distribution functions are substituted by their empirical counterparts.

Parametric estimation - One-step estimation Within the one-step estimation the parameters $\theta, p_1, \ldots, p_n$ are estimated in *one* step by maximum-likelihood (ML) methods. Under the usual regularity conditions (see e.g. Casella and Berger (2002), p. 516) for asymptotic ML-theory, the functional estimator $\widehat{\lambda} = \lambda(\widehat{\theta})$ represents an ML estimate of λ. Hence, consistency and asymptotic normality are given, i.e.,

$$\sqrt{m}\left(\widehat{\lambda}_m - \lambda\right) \xrightarrow{d} Z \sim N\left(0, \sigma^2\right) \quad \text{as } m \to \infty,$$

where σ^2 denotes the asymptotic variance corresponding to $\widehat{\lambda}_m$. In case $\lambda(\cdot)$ is a continuous differentiable function in $\theta = (\theta_1, \ldots, \theta_l)$, the standard error σ can be approximated by the usual Delta Method (see Casella and Berger (2002), Section 5.5.4), in particular

$$\widehat{\sigma}^2 = \sum_{i=1}^{l} \left(\frac{\partial \lambda}{\partial \theta_i} \left(\widehat{\theta}_i \right) \right)^2 \sigma^2_{\widehat{\theta}_i}, \tag{4.51}$$

where $\sigma_{\hat{\theta}_i},\ i = 1, \ldots, l$, denotes the standard error of the corresponding ML estimate of $\theta_i,\ i = 1, \ldots, l$.

Parametric estimation - Two-step estimation One of the advantages of the copula concept is the preservation of good statistical estimation properties while estimating the parameters $\theta, p_1, \ldots, p_n$ in *two* steps. Here we assume that the copula parameter θ belongs only to the copula function C_θ and is independent of the marginal distribution functions. We first estimate the marginal parameters $p_1, \ldots, p_n$ and then fit the marginal distribution functions $F_1, \ldots, F_n$ to the observed realization $x^{(1)}, \ldots, x^{(m)}$ of the random sample $X^{(1)}, \ldots, X^{(m)}$. In a second step the copula parameter θ is estimated from approximate realizations of the corresponding copula, i.e., $\hat{u}^{(1)}, \ldots, \hat{u}^{(m)}$ with $\hat{u}_j^{(i)} := F_{\hat{p}_j}(x_j^{(i)}),\ j = 1, \ldots, n,\ i = 1, \ldots, m$. Denote $\hat{u} := (\hat{u}^{(1)}, \ldots, \hat{u}^{(m)})$ as the matrix of approximate copula observations. Joe (1997), Chapter 10, provides the asymptotic distribution and asymptotic covariance matrix for the latter estimator. Note, this method is also referred to as pseudo-maximum likelihood. In Section 4.11 our simulation study reveals that the relative efficiency, which is defined as the ratio between the MSE of the two-step ML estimator and the MSE of the one-step ML estimator, is close to 1. Thus, there is not much difference between estimating the TDC via two steps or via one step. However, the two-step approach simplifies the numerics of the parameter estimation a lot. Especially for high-dimensional data sets, the TDC estimation in two steps becomes a relief.

Semi-parametric estimation - Two-step estimation To avoid misidentifications of the univariate marginal distributions, the marginal distribution functions $F_1, \ldots, F_n$ are now estimated by their empirical distribution functions (nonparametric estimation). In a second step the copula parameter θ is estimated from the approximate copula realizations $\hat{u}$ via an ML-estimator, i.e., $\hat{\theta} = \hat{\theta}(\hat{u})$. As shown by Genest, Ghoudi, and Rivest

(1995) and Shih and Louis (1995), the resulting estimator is consistent and asymptotically normal. For the asymptotic covariance matrix and a corresponding estimate we refer to Genest, Ghoudi, and Rivest (1995). In case $\lambda(\cdot)$ represents a smooth function, the estimator $\hat{\lambda} = \lambda(\hat{\theta})$ turns out to be also consistent and asymptotically normal. The latter two-step estimator is favorably used in some recent empirical investigations; see e.g. Ané and Kharoubi (2001), and Junker and May (2002).

On the other hand, marginal distributions are often better known than the dependence function. For instance, it is appropriate to assume normal distributed margins in many financial models whereas the dependence structure cannot be clearly specified. Therefore, the following semi-parametric method specifies the marginal distributions via a parametric procedure in a first step, and estimates the related copula via some nonparametric estimator in a second step. For instance, at a first stage, we may utilize an ML-estimator for the parameters $p_1, \ldots, p_n$ of the marginal distribution functions $F_1, \ldots, F_n$ and, at a second stage, the copula C is estimated via $\hat{C}_{\hat{p_1},\ldots,\hat{p_n}} = F_m(F_{\hat{p}_1}^{-1}, \ldots, F_{\hat{p}_n}^{-1})$, where F_m denotes the multivariate empirical distribution function of F. Under suitable regularity conditions the process

$$\sqrt{m}(\hat{C}_{\hat{p_1},\ldots,\hat{p_n}} - C) = \sqrt{m}(F_m(F_{\hat{p}_1}^{-1}, \ldots, F_{\hat{p}_n}^{-1}) - F(F_1^{-1}, \ldots, F_n^{-1})$$

converges to a Gaussian limit (cf. Fermanian, Radulović, and Wegkamp (2002)). The TDC is now estimated via the limiting relation (2.14).

4.10.2 Estimation from the distribution's tail or extreme samples

Threshold method

Like in Section 4.10.1 we choose a parametric copula family allowing for tail dependence, and estimate the TDC via a function of the inferred distribution parameters. However, in the present section we assume that the underlying model is only appropriate for the joint extremes and not for the entire data sample. Hence, we choose lower and upper n-dimensional thresholds and investigate only data below the threshold (th_L) and above the threshold (th_U), respectively. The lower and upper TDC are estimated from these data. In case of radial symmetry where $\lambda_L = \lambda_U$ we only have to determine one threshold th, and use all data points below th and above $1 - th$ at once.

The simulation study (cf. Section 4.11) uses a truncated Gumbel copula (cf. Table 3.2) to model upper tail-dependence. The Gumbel copula seems to be a natural choice as it is the copula of an extreme value distribution (see Joe (1997)). In case of radial symmetry we mirror the lower joint extreme values of the copula via $(u_1, u_2) \to (1 - u_1, 1 - u_2)$ to the upper right corner. Finally, the threshold is chosen such that $2\sqrt{m}$ observations are in the upper right corner $[th, 1] \times [th, 1]$.

Nonparametric methods

In the present section we drop any assumptions on the marginal distributions or the copula function, and provide various nonparametric estimators for the lower and the upper

TDC, respectively. Nonparametric estimation methods prevent wrong specifications of the underlying distribution or copula and thus avoid misidentifications of the upper or lower TDC. Here we consider a bivariate random sample $(X_1^{(1)}, X_2^{(1)}), \ldots, (X_1^{(m)}, X_2^{(m)})$ of iid bivariate random vectors with distribution function F and copula C.

The following nonparametric estimators have already been introduced in Sections 4.2 and 4.9.2:

$$\hat{\lambda}_{U,m} = \frac{1}{k} \cdot \sum_{j=1}^{m} \mathbf{1}_{\{R_{m1}^{(j)} > m-k \text{ and } R_{m2}^{(j)} > m-k\}}, \, \hat{\lambda}_{L,m} = \frac{1}{k} \cdot \sum_{j=1}^{m} \mathbf{1}_{\{R_{m1}^{(j)} \leq k \text{ and } R_{m2}^{(j)} \leq k\}},$$

$$\hat{\lambda}_{U,m}^{EVT} = 2 - \frac{1}{k} \cdot \sum_{j=1}^{m} \mathbf{1}_{\{R_{m1}^{(j)} > m-k \text{ or } R_{m2}^{(j)} > m-k\}}, \; 1 \leq k \leq m,$$

with $k = k(m) \to \infty$ and $k/m \to 0$ as $m \to \infty$. Asymptotic normality and strong consistency have been discussed in Sections 4.5 and 4.6.

Another nonparametric estimator is motivated by the limiting relation (2.14). If the diagonal section $v \mapsto C(v,v)$ is differentiable on the interval $(1-\varepsilon, 1)$ for some $\varepsilon > 0$, then

$$\lim_{v \to 1^-} \frac{1 - C(v,v)}{1-v} = \lim_{v \to 1^-} \frac{dC(v,v)}{dv} = \lim_{v \to 1^-} \frac{\log C(v,v)}{\log v}. \tag{4.52}$$

Therefore an estimator for the upper TDC is provided by

$$\begin{aligned} \hat{\lambda}_{U,m}^{LOG} &= 2 - \frac{\log C_m(\frac{m-k}{m}, \frac{m-k}{m})}{\log \frac{m-k}{m}} \\ &= 2 - \frac{\log \frac{1}{m} \cdot \sum_{j=1}^{m} \mathbf{1}_{\{R_{m1}^{(j)} \leq m-k \text{ and } R_{m2}^{(j)} \leq m-k\}}}{\log \frac{m-k}{m}}, \qquad 1 \leq k < m, \end{aligned} \tag{4.53}$$

with $k = k(m) \to \infty$, $k/m \to 0$ as $m \to \infty$ and C_m being the empirical copula. If the sample data are either comonotonic or stochastically independent, the above estimator turns out to behaved well for all k because in the first case $C(v,v) = v$ and thus

$$\lambda_U = 2 - \frac{\log C(v,v)}{\log v} = 2 - \frac{\log v}{\log v} = 1 \text{ for all } v \in (0,1), \tag{4.54}$$

whereas in the second case $C(v,v) = v^2$ and hence

$$\lambda_U = 2 - \frac{\log C(v,v)}{\log v} = 2 - \frac{\log v^2}{\log v} = 2 - 2 \cdot \frac{\log v}{\log v} = 0 \text{ for all } v \in (0,1). \tag{4.55}$$

This means that the TDC is given exactly at each point on the logarithmic diagonal section of the copula if the diagonal section has a power law which indeed is the case for comonotonicity and independence. This property does not hold for the estimators given in the first part of the present section. Note that if $C(v,v) = v^\alpha, \; \alpha \in [1,2]$, for all $v \in (1-\varepsilon, 1)$ then $\lambda_U = 2 - \alpha$.

A similar estimator for the lower TDC is motivated by Proposition 2.2.5. It can be shown that

$$\lim_{v\to 0^+} \frac{dC(v,v)}{dv} = 2 - \lim_{v\to 0^+} \frac{\log(1-2v+C(v,v))}{\log(1-v)}, \tag{4.56}$$

if the diagonal section $v \mapsto C(v,v)$ is differentiable on the interval $(0,\varepsilon)$ for some $\varepsilon > 0$. Therefore an estimator for the lower TDC λ_L is provided by

$$\begin{aligned}\widehat{\lambda}^{LOG}_{L,m} &= 2 - \frac{\log C_m\left(\left(\frac{k}{m},1\right] \times \left(\frac{k}{m},1\right]\right)}{\log\frac{m-k}{m}} \\ &= 2 - \frac{\log \frac{1}{m}\cdot \sum_{j=1}^m \mathbf{1}_{\{R^{(j)}_{m1}>k \text{ and } R^{(j)}_{m2}>k\}}}{\log\frac{m-k}{m}}, \qquad 1 \le k < m, \end{aligned} \tag{4.57}$$

with $k = k(m) \to \infty$ and $k/m \to 0$ as $m \to \infty$. Also if $1-2v+C(v,v) = (1-v)^\alpha$, $\alpha \in [1,2]$, for all $v \in (0,\varepsilon)$ then $\lambda_L = 2-\alpha$. Moreover, for the comonotonic copula ($\alpha = 1$) and the product copula ($\alpha = 2$) the latter estimator fulfills the same stability properties as derived for the log-estimator in (4.54) and (4.55).

However, the finite sample properties of the log-estimators $\widehat{\lambda}^{LOG}_{U,m}$ and the estimator $\widehat{\lambda}_{U,m}$ defined in (4.48) are not significantly different in the case of tail independence as illustrated in Figure 4.5. Here we plot the empirical mean and the corresponding confidence interval of 1000 upper TDC estimates for various k based on a bivariate (standard) normal distribution with correlation 0.5 (cf. Theorem 3.2.7). Note that the smoothing effect of the logarithm causes a decreasing bias.

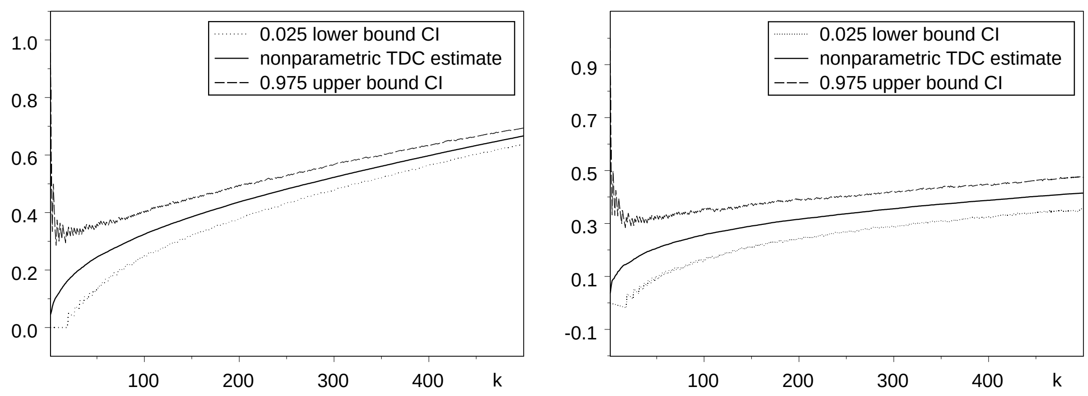


Figure 4.5: Plots of empirical mean and the corresponding confidence interval (CI) of 1000 upper TDC estimates $\widehat{\lambda}_{U,m}$ defined in (4.48) (left plot) and upper TDC estimates $\widehat{\lambda}^{LOG}_{U,m}$ defined in (4.53) (right plot) for various k based on a bivariate (standard) normal distribution with correlation 0.5.

Regarding the statistical properties of the log-estimators (4.53) and (4.57), strong consistency holds if $k/\log\log m \to \infty$ as $m \to \infty$. This can be immediately deduced

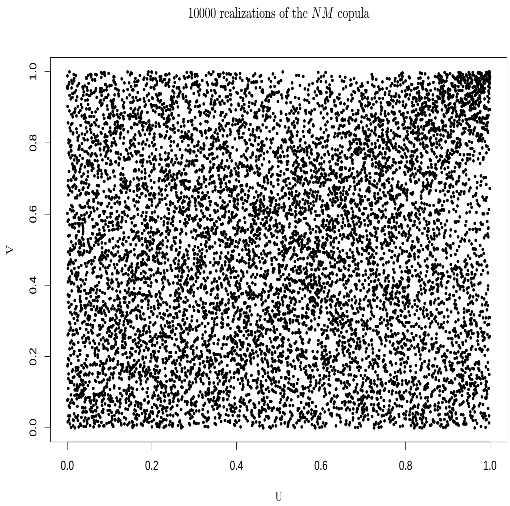


Figure 4.6: Scatterplot of simulated data from a normal mixture distribution NM.

from the strong consistency result for the nonparametric estimators $\hat{\lambda}_{U,m}$ and $\hat{\lambda}_{L,m}$ (cf. Section 4.6).

4.10.3 Pitfalls

We would like to point out that starting from a finite iid random sample $X,\ X_1,\ldots,X_n$, it is nearly impossible to determine definitely whether X is tail dependent or not. Even for large sample sizes we can find certain distribution models which produce finite samples pretending tail dependence (similar problems occur in the context of tail index estimations). For example Figure 4.6 shows the scatterplot of realizations from a copula corresponding to a mixture distribution of two different bivariate Gaussian distributions, namely

$$NM = \frac{1}{2}N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix}\right) + \frac{1}{2}N\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1.44 & 0.72 \\ 0.72 & 1.44 \end{bmatrix}\right).$$

At the first glance the scatterplot reveals upper tail-dependence although the NM distribution is tail independent. This example clearly shows that it is important to be careful by inferring tail dependence from a finite random sample. The best way to prevent misidentifications is the application of several estimators, tests or plots to the same data set.

Moreover, the use of a distribution model with parametric margins instead of empirical margins contains also the risk of misidentification and may lead to possible copula mapping-error. This results in a misinterpretation of the dependence structure. For example, we simulated 3000 realizations from a bivariate distribution function $H(x,y) = C_{G,\vartheta=2}\left(t_{\nu=3}(x), t_{\nu=3}(y)\right)$. Here t_ν denotes the univariate standard t-distribution with ν degrees of freedom and C_G denotes the Gumbel copula with

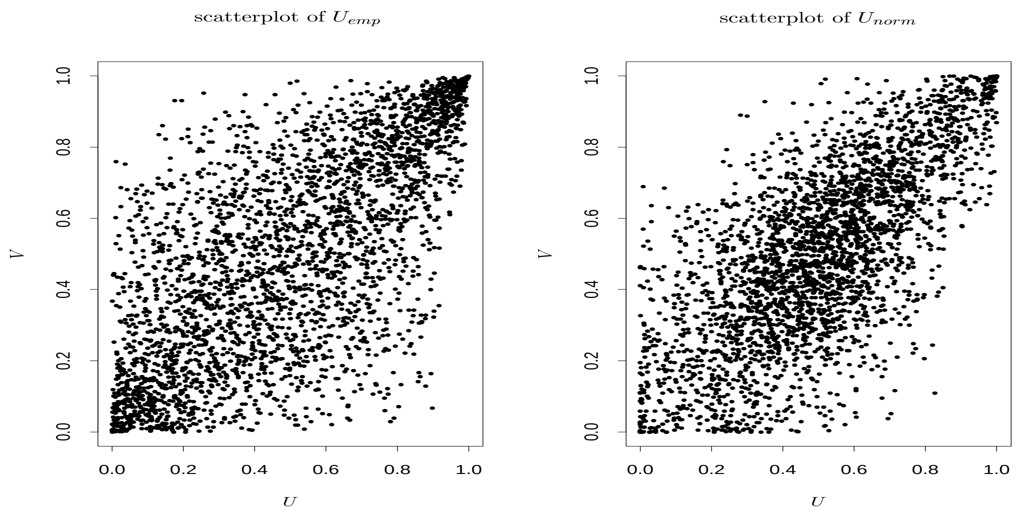


Figure 4.7: Comparison of copula scatterplots corresponding to different marginal mappings. The right plot contains the misspecified margins.

parameter $\nu = 2$. Then we transform the one-dimensional margins to uniform distributions using their empirical distributions. The resulting 2-dimensional approximate copula sample is denoted by U_{emp}. The latter is compared to the 2-dimensional approximate copula sample U_{norm} which is realized by transforming the margins via a fitted univariate normal distributions, i.e., $U_{norm} = \left\{ \left(\widehat{\Phi}_X (x_i), \widehat{\Phi}_Y (y_i) \right) \right\}$. Figure 4.7 clearly shows that U_{emp} reflects the simulated dependence of a Gumbel copula whereas U_{norm} has nearly lost all appearance of upper tail-dependence and shows evidence of a sort of "linear" dependence. We remark that the hypothesis of uniform margins of U_{norm} is usually rejected by KS-test statistics. Summarizing the latter, we showed that a wrong marginal fit can cause dramatic misinterpretation of the underlying dependence structure.

4.11 Simulation study

In order to compare the properties and characteristics of the TDC estimators introduced in the last two sections, we run an extensive simulation study. Each simulated data set consists of 1000 independent copies of m realizations obtained from a random sample $X^{(1)}, \ldots, X^{(m)}$ following one distribution out of four. Three different sample sizes m are considered for each data set, in particular $m = 250, 1000, 5000$. The four different distributions under consideration are denoted by $\mathcal{H}$, $\mathcal{T}$, $\mathcal{F}$ and $\mathcal{G}$. For example, the data set $S_{\mathcal{H}}^{250}$ is defined by

$$S_{\mathcal{H}}^{250} := \Big\{ \big\{ X^{(i,j)}(\omega) : X^{(i,j)} \sim \mathcal{H}(\lambda = 1, \alpha = 1, \rho = 0.5), \\ j = 1, \ldots, 250 \big\} : i = 1, \ldots, 1000 \Big\},$$

where $\mathcal{H}(\lambda = 1, \alpha = 1, \rho = 0.5)$ denotes the standard bivariate symmetric generalized-hyperbolic distribution with correlation coefficient $\rho = 0.5$ as we will define in Section 5.1.1. Distribution $\mathcal{T}$ corresponds to the Student's t-distribution with $\nu = 1.5$ degrees of freedom and $\rho = 0.5$. Further, another elliptical distribution $\mathcal{F}$ is characterized by the following random variate R_2 and scaling matrix

$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix},$$

$$R_2 \sim \left(\frac{1}{2} \cdot \mathrm{LN}(\mu = 1, \sigma = 0.01) + \frac{1}{2} \cdot \mathrm{GPD}(\xi = \frac{2}{3}, \beta = 1)\right),$$

i.e., R_2 is the mixture of a lognormal distribution and a generalized Pareto distribution.

The last distribution function $\mathcal{G}$ is defined by

$$\mathcal{G}(x_1, x_2) := C_{\vartheta,\gamma}\left(\Phi(x_1), \Phi(x_2)\right),$$

where Φ is the univariate standard normal distribution function and $C_{\vartheta,\gamma}$ is an Archimedean copula (cf. Section 3.4) achieved by a transformed Frank copula with generator

$$\varphi_{\vartheta,\gamma}(t) = (\varphi_{\text{Frank}}(t))^\gamma = \left(-\log \frac{e^{-\vartheta t} - 1}{e^{-\vartheta} - 1}\right)^\gamma, \qquad \vartheta \neq 0, \gamma < 1,$$

introduced in Junker and May (2002). Here we set $\vartheta = 1.56$ and $\gamma = -0.76$.

Note that $\mathcal{H}$ represents an elliptically contoured distribution which is tail independent. The corresponding realizations are used to investigate the stability of the utilized estimation procedures under the absence of tail dependence. In contrast to that, $\mathcal{T}$ and $\mathcal{F}$ are elliptically contoured distributions possessing tail dependence (see Section 3.2.2). The corresponding parameters are: $\rho = \Sigma_{12}/\sqrt{\Sigma_{11}\Sigma_{22}} = 0.5$ and tail index $\alpha = -1.5$, and the TDC is obtained via the relation

$$\lambda = 2 \cdot \bar{t}_{1-\alpha}\left(\sqrt{1-\alpha} \cdot \sqrt{\frac{1-\rho}{1+\rho}}\right). \tag{4.58}$$

Observe that the copula $C_{\vartheta,\gamma}$ is lower tail-independent but upper tail-dependent, i.e., $\lambda_L^{\mathcal{G}} = 0$ and $\lambda_U^{\mathcal{G}} = 2 - 2^{\frac{1}{\gamma}}$. In order to provide sufficient comparability of the estimation results, the parameterizations of the four models $\mathcal{H}, \mathcal{T}, \mathcal{F}$, and $\mathcal{G}$ are chosen such that $\lambda_L^{\mathcal{T}} = \lambda_U^{\mathcal{T}} = \lambda_L^{\mathcal{F}} = \lambda_U^{\mathcal{F}} = \lambda_U^{\mathcal{G}} = 0.4406$ and $\tau^{\mathcal{H}} = \tau^{\mathcal{T}} = \tau^{\mathcal{F}} = \tau^{\mathcal{G}} = 0.\overline{3}$. Recall from Section 3.3.1 that Kendall's τ for elliptical copulae can be obtained via the relation

$$\tau = \frac{2}{\pi} \cdot \arcsin(\rho). \tag{4.59}$$

Figure 4.8 illustrates the different tail-behavior of the $\mathcal{H}, \mathcal{T}, \mathcal{F}$ and $\mathcal{G}$ distribution by presenting the scatterplots of the respective simulated data-sample with sample size $m = 5000$ together with the corresponding empirical copula-density. Regarding the copula-mapping we utilize marginal empirical distribution functions.

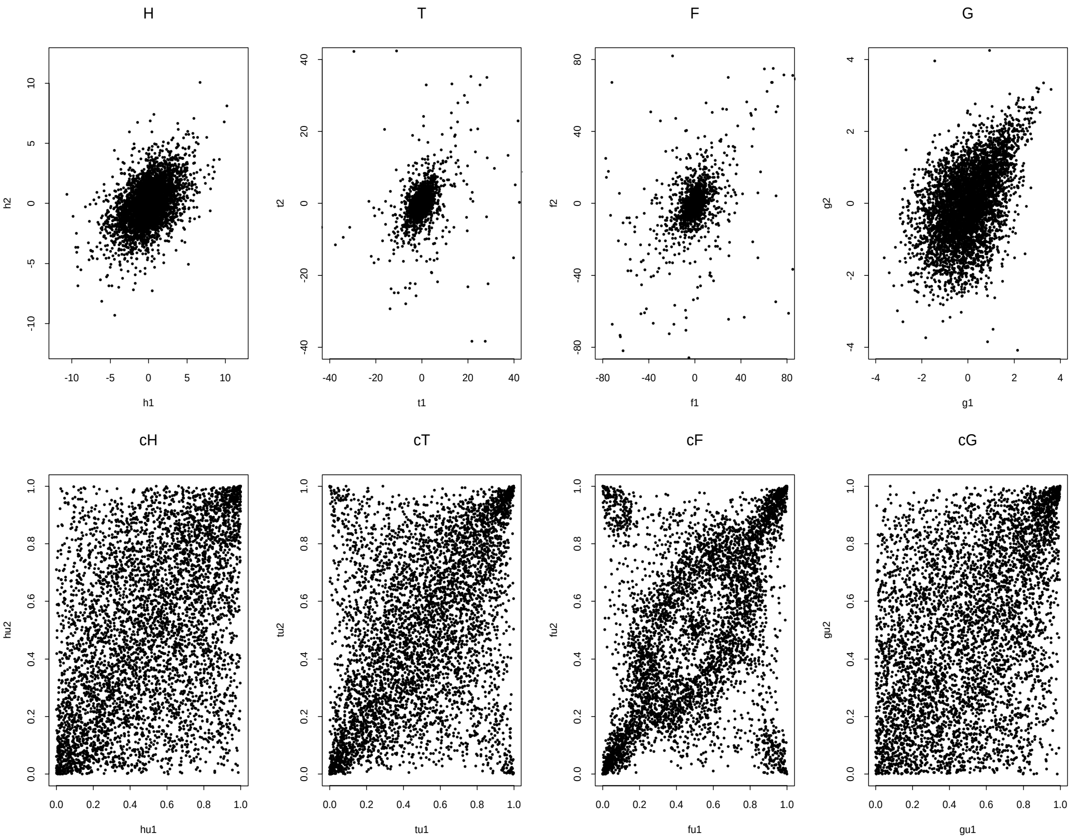


Figure 4.8: Scatterplots of $m = 5000$ simulated data following either the $\mathcal{H}, \mathcal{T}, \mathcal{F}$ or $\mathcal{G}$ distribution (first row) and corresponding empirical copula-densities (second row).

4.11.1 Pre-simulation

Referring to Section 4.10.1, in the present pre-simulation we begin with the comparison of the statistical properties of the two-step estimation (empirical margins and ML estimation of $\theta(\widehat{u})$) with the one-step estimation (ML estimation of θ) for the $S_{\mathcal{T}}^{1000}$ data set with parameters $\nu = 1.5$ and $\rho = 0.5$. The one-step ML estimator is denoted by $\widehat{\theta}_{\mathcal{T},ML} := (\widehat{\nu}_{ML}, \widehat{\rho}_{ML})$ and the two-step estimator is denoted by $\widehat{\theta}_{\mathcal{T},\widehat{u}} := (\widehat{\nu}, \widehat{\rho})$. Comparing the empirical means, the medians, and the standard deviations

$$\begin{aligned} \overline{\widehat{\theta}_{\mathcal{T},ML}} &= (1.5364, 0.5000)\,, & \overline{\widehat{\theta}_{\mathcal{T},\widehat{u}}} &= (1.5388, 0.5004)\,, \\ \mu_{0.5}\big(\widehat{\theta}_{\mathcal{T},ML}\big) &= (1.5003, 0.5001)\,, & \mu_{0.5}\big(\widehat{\theta}_{\mathcal{T},\widehat{u}}\big) &= (1.5242, 0.5005) \\ \sigma\big(\widehat{\theta}_{\mathcal{T},ML}\big) &= (0.3562, 0.0293)\,, & \sigma\big(\widehat{\theta}_{\mathcal{T},\widehat{u}}\big) &= (0.1447, 0.0316)\,, \end{aligned}$$

we conclude that there is almost no difference regarding the estimation of ρ. However, the standard deviations of $\widehat{\nu}_{ML}$ and $\widehat{\nu}$ differ significantly as a result of upper outliers of $\widehat{\nu}_{ML}$. Another statistical measure is provided by the ratio of mean squared errors

$$MSE\big(\widehat{\theta}_{\mathcal{T},ML}\big)/MSE\big(\widehat{\theta}_{\mathcal{T},\widehat{u}}\big) = (5.71, 0.86)\,.$$

The ratio confirms the worse estimation of $\widehat{\nu}_{ML}$ compared to $\widehat{\nu}$ and the satisfying estimation of the correlation coefficient. Therefore the two-step estimation has much to recommend on.

We continue with a comparison between the estimator for the correlation coefficient of bivariate elliptical random vectors introduced in Section 4.9.1 (and denoted by $\widehat{\rho}_\Gamma$) and the estimator for the correlation coefficient introduced in Lindskog, McNeil, and Schmock (2001) (denoted by $\widehat{\rho}_\tau$) which is based on Kendall's tau τ (see also Section 3.3.1). This is of interest because the TDC for elliptical distributions heavily depends on the estimate of the correlation coefficient according to relation (4.38). For the estimation of the correlation coefficient from $S_{\mathcal{T}}^{1000}$, Figure 4.9 reveals that the theoretical biases

$$\big(\overline{\widehat{\rho}_\tau}, \overline{\widehat{\rho}_\Gamma}\big) = (0.5008, 0.4988)$$

are negligible and the standard deviations

$$\left(\sigma\left(\widehat{\rho}_\tau\right), \sigma\left(\widehat{\rho}_\Gamma\right)\right) = (0.0326, 0.0481)$$

are of comparable size. The mean square error ratio $MSE\left(\widehat{\rho}_\Gamma\right)/MSE\left(\widehat{\rho}_\tau\right) = 2.18$ indicate a favour for $\widehat{\rho}_\tau$ because of its lower volatility. However, in higher dimensions, the estimate $\widehat{\rho}_\Gamma$ immediately yields a correlation matrix whereas the estimator $\widehat{\rho}_\tau$ has to be adjusted in order to guarantee positive semi-definiteness of the correlation matrix.

According to Section 4.9.2, two possibilities are available to estimate the tail index α of an elliptical distributed random vector X. Either we may utilize realizations of $\|X\|_2$ or approximate realizations of R_n^*. The impact of both methods on the tail index estimation is compared in Figure 4.10 via two average mean excess plots corresponding to realization of $\|X\|_2$ and R_n^* of the data set $S_{\mathcal{T}}^{1000}$. It turns out that the slope of the fitted linear regression line coincides for both procedures. Hence, the tail index

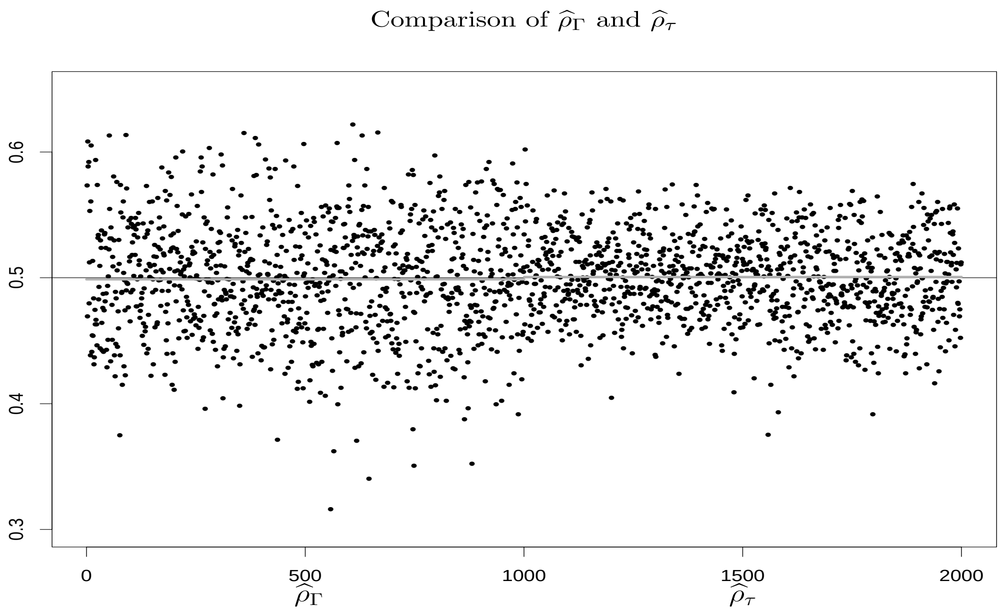


Figure 4.9: Comparison of 1000 estimates of $\widehat{\rho}_\Gamma$ (left half of the plot) and 1000 estimates of $\widehat{\rho}_\tau$ (right half of the plot). The black line relates to the theoretical value.

estimation is quite similar. However, note that in general $\|X\|_2$ and R_n^* have different distribution functions (Figure 4.11 shows the corresponding average empirical survival distributions of S_T^{1000}).

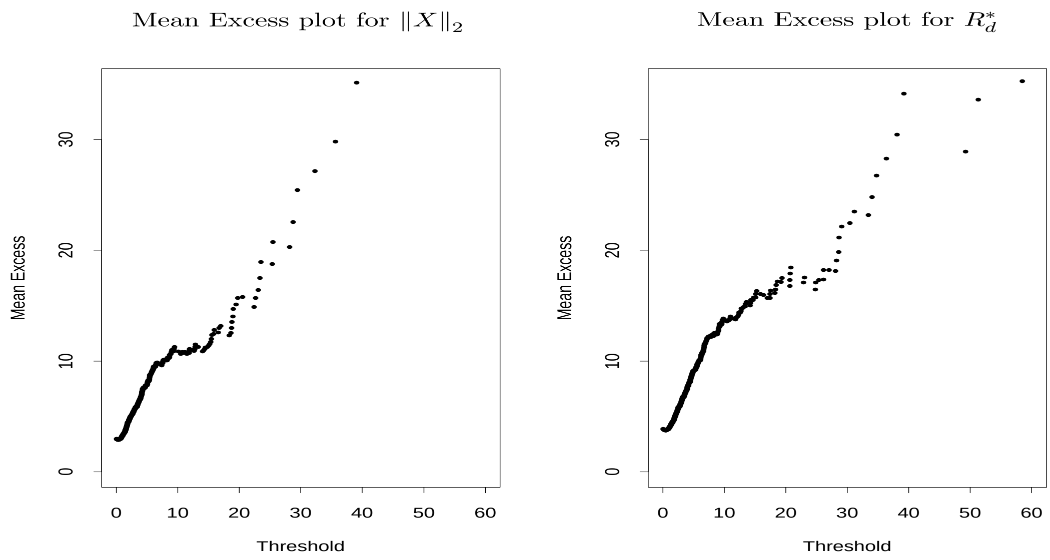


Figure 4.10: Mean Excess Plots for a realization of $\|X\|_2$ and R_d^*.

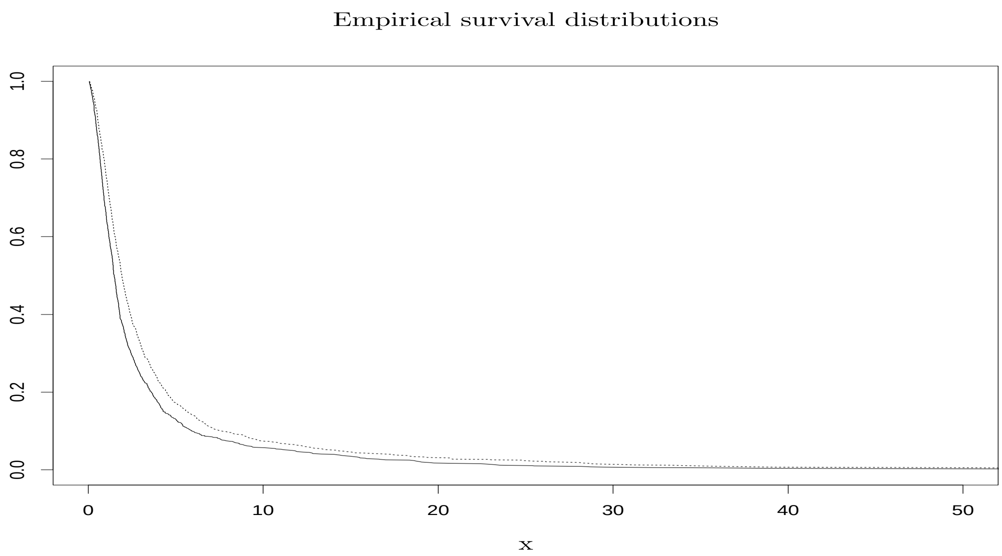


Figure 4.11: Average empirical survival distributions of $\|X\|_2$ (dotted line) and R_n^* (solid line) corresponding to the data set $S_{\mathcal{I}}^{1000}$.

4.11.2 Simulation and estimation

In the present section we again restrict ourselves to the simulation and estimation of the upper TDC λ_U. For convenience we drop the subscript U if it is clear from the context. For the comparison of the different estimation methods we calculate the sample mean $\overline{\overline{\lambda}}_m$, the sample standard deviation $\sigma(\widehat{\lambda}_m)$, and the root mean square-error (RMSE)

$$\mathrm{RMSE}\left(\widehat{\lambda}_m\right) := \sqrt{\frac{1}{m} \cdot \sum_{i=1}^{m} \left(\widehat{\lambda}_{m,i} - \lambda\right)^2}. \tag{4.60}$$

The RMSE allows to observe the bias-variance trade-off between different sample sizes and estimation procedures. Further we consider another statistical figure given by

$$\mathrm{MESE}\left(\widehat{\lambda}_m\right) := \frac{\mathrm{RMSE}\left(\widehat{\lambda}_m\right)}{\sigma\left(\widehat{\lambda}_m\right)} - 1 \tag{4.61}$$

which measures the distance between the standard error $\sigma(\widehat{\lambda}_m)$ and the bias due to model misidentification. The abbreviation MESE refers to mean error - standard error. Note that $\mathrm{MESE}(\widehat{\lambda}_m) \to 0$ as $n \to \infty$ if the estimator of λ is asymptotically unbiased, since in that case $\mathrm{RMSE}(\widehat{\lambda}_m) \to \sigma(\widehat{\lambda}_m)$ as $m \to \infty$. For every asymptotically biased estimation the RMSE is bounded away from zero and therefore the MESE-limit diverges to infinity.

Estimation from the entire sample

Specifying a parametric copula The present section relates to Section 4.10.1. Regarding the elliptical data sets $S_{\mathcal{H}}$, $S_{\mathcal{T}}$, and $S_{\mathcal{F}}$, we fit a t-copula since this seems to be a realistic choice by glancing only at the scatterplots in Figure 4.8. However, the empirical copula density of these data sets indicate that the t-copula may not be suitable for $S_{\mathcal{F}}$. Although the true copula-model fails in two cases, we estimate the upper TDC via the relation (4.58). In Figure 4.12, 3×1000 consecutive TDC estimates corresponding to $S_{\mathcal{T}}$ are presented.

In the following we will plot the estimates for the data set $S_{\mathcal{T}}$ under different estimation procedures. For a better comparability we fix the range of the plots. Regarding the data set $S_{\mathcal{G}}$, we fit a Gumbel copula which seems to be appropriate according to Figure 4.8, and estimate the TDC via the relation $\widehat{\lambda}^{\mathcal{G}} = 2 - 2^{\frac{1}{\widehat{\gamma}}}$. This relationship was established in Table 3.2 for the Gumbel copula.

The estimation results for all data sets are summarized in Tables 4.6 and 4.7.

Estimation from the distribution's tail or extreme samples

The threshold method For the threshold approach, introduced in Section 4.10.2, the statistical figures corresponding to the data sets $S_{\mathcal{H}}$, $S_{\mathcal{T}}$,$S_{\mathcal{F}}$ and $S_{\mathcal{G}}$ are provided in Table 4.6. Figure 4.13 again gives a graphical survey for the data-set $S_{\mathcal{T}}$.

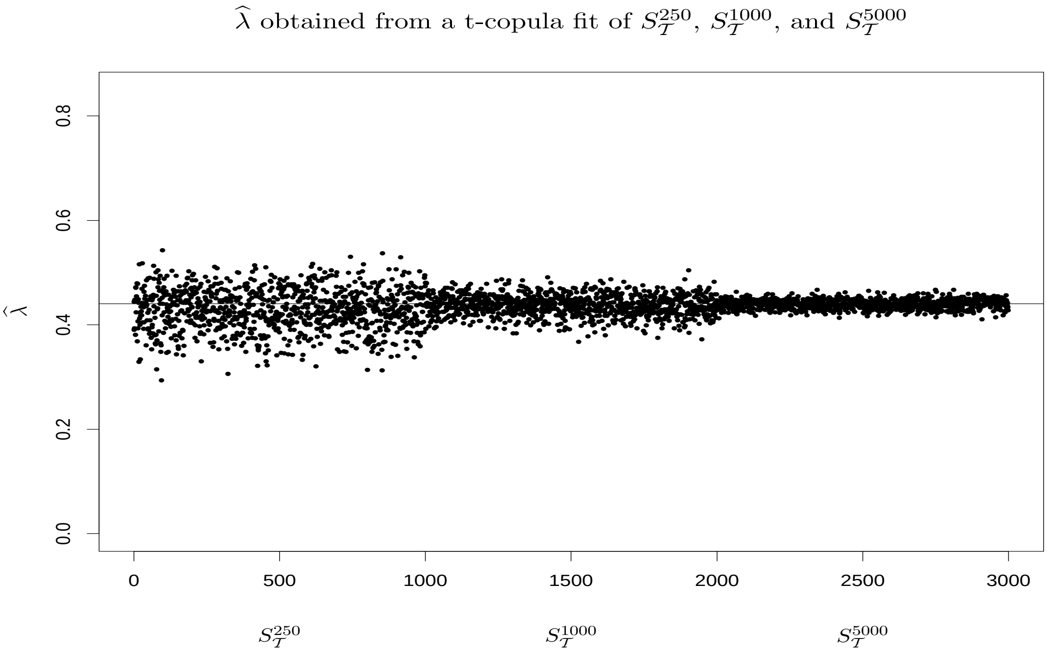


Figure 4.12: Consecutive estimates of λ from $S_{\mathcal{T}}^{250}$, $S_{\mathcal{T}}^{1000}$, and $S_{\mathcal{T}}^{5000}$ utilizing a t-copula and empirical margins, cf. Section 4.10.1 (the black line is the true value).

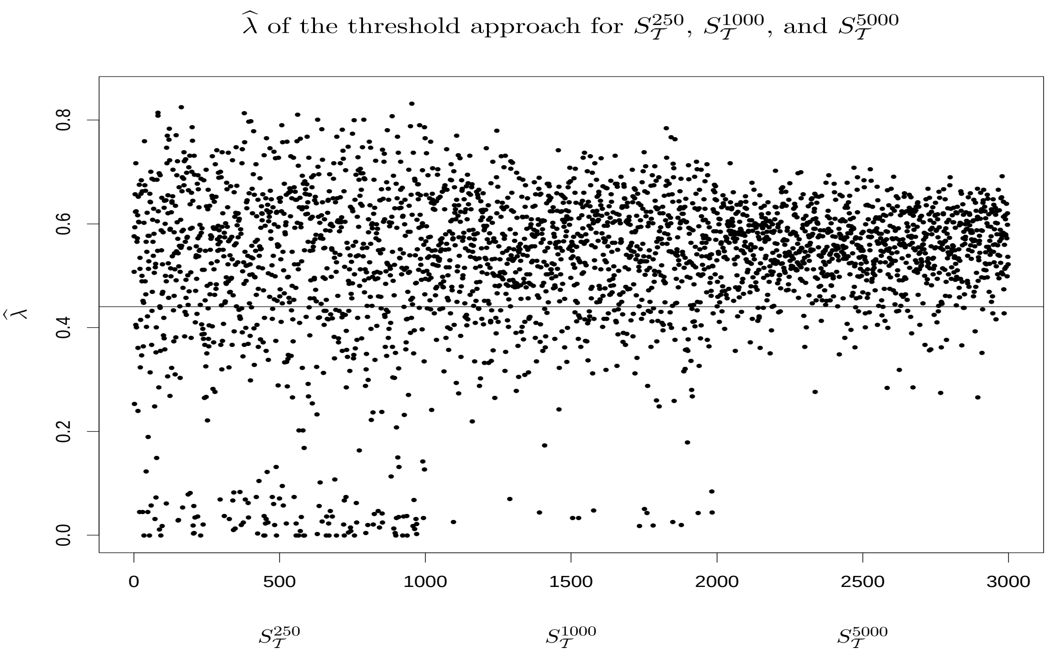


Figure 4.13: Consecutive estimates of λ from $S_{\mathcal{T}}^{250}$, $S_{\mathcal{T}}^{1000}$, and $S_{\mathcal{T}}^{5000}$ utilizing the threshold approach, cf. Section 4.10.2 (the black line is the true value).

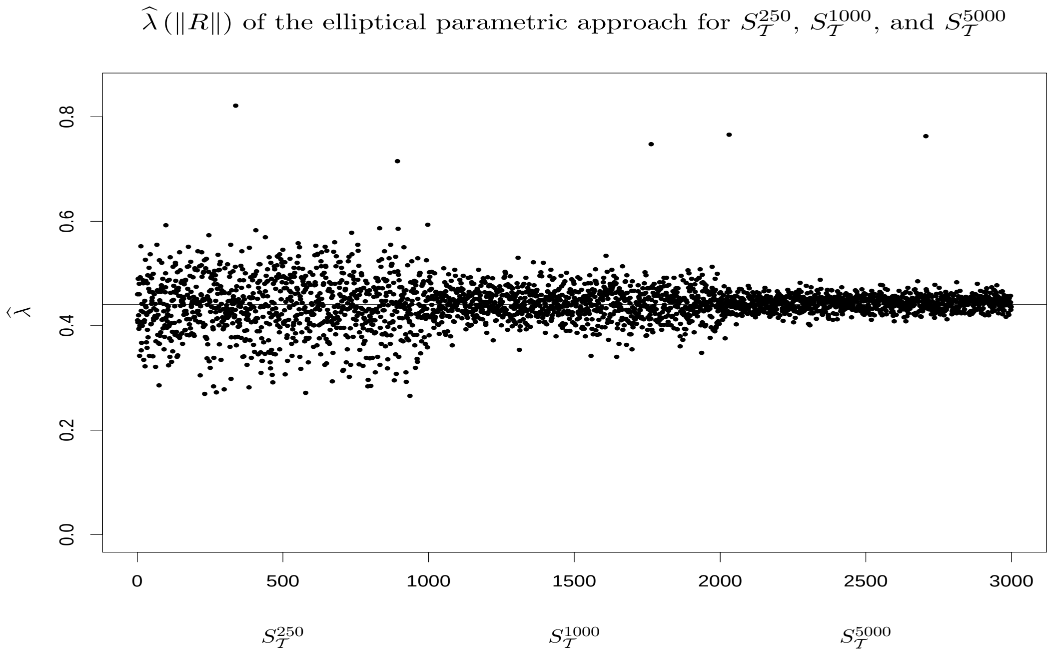


Figure 4.14: Consecutive estimates of λ from $S_{\mathcal{T}}^{250}$, $S_{\mathcal{T}}^{1000}$, and $S_{\mathcal{T}}^{5000}$ utilizing the parametric approach for elliptically contoured distributions stated in Section 4.9.1, where the tail index is estimated via R_n^* (the black line is the true value).

Parametric method for elliptical distributions Regarding the data sets $S_{\mathcal{H}}, S_{\mathcal{T}}$ and $S_{\mathcal{F}}$, we have estimated the correlation matrix and the tail index via the estimation method introduced in Section 4.9.1. The estimates for the TDC are derived via relation (4.58) and are visualized in Figure 4.14. Additionally, for the data-set $S_{\mathcal{T}}$, we estimate the tail index via $\|X\|_2$ and the correlation coefficient via Kendall's τ. The results are summarized in Figure 4.15.

Obviously the data of $S_{\mathcal{G}}$ are not elliptically distributed. Therefore we skip the latter estimation method for $S_{\mathcal{G}}$. Again, the corresponding statistical figures are presented in Table 4.6 at the end of the present section.

Nonparametric approach Regarding the data sets $S_{\mathcal{H}}, S_{\mathcal{T}}, S_{\mathcal{F}}$ and $S_{\mathcal{G}}$, the statistical figures of the nonparametric estimator (4.48) and the log-estimator (4.53), introduced in Section 4.10.2, are provided in Table 4.7. For the data set $S_{\mathcal{T}}$, Figure 4.16 gives a graphical survey corresponding to the estimator (4.48).

The plateaus are chosen by a heuristic plateau-finding algorithm that first smooths the plateau-plot by a simple box kernel with 'bandwidth' bd, i.e., the mean of $bd+1$ successive points of the plateau vector $(\lambda(1), \ldots, \lambda(k), \ldots, \lambda(m))$ lead to a new point of the smoothed plateau $\Lambda = (\lambda(1), \ldots, \lambda(m-bd))$. We take $bd = \lfloor 0.005 \cdot n \rfloor$. In a second step, a plateau of length $l = \lfloor \sqrt{m-bd} \rfloor$ is defined as the vector

$$(\lambda(i), \lambda(i+1), ..., \lambda(i+l-1)), \quad i \in \{1, ..., m-bd\}.$$

The algorithm take the first plateau $k \in \{1, ..., m-bd-l+1\}$ whose elements fulfill

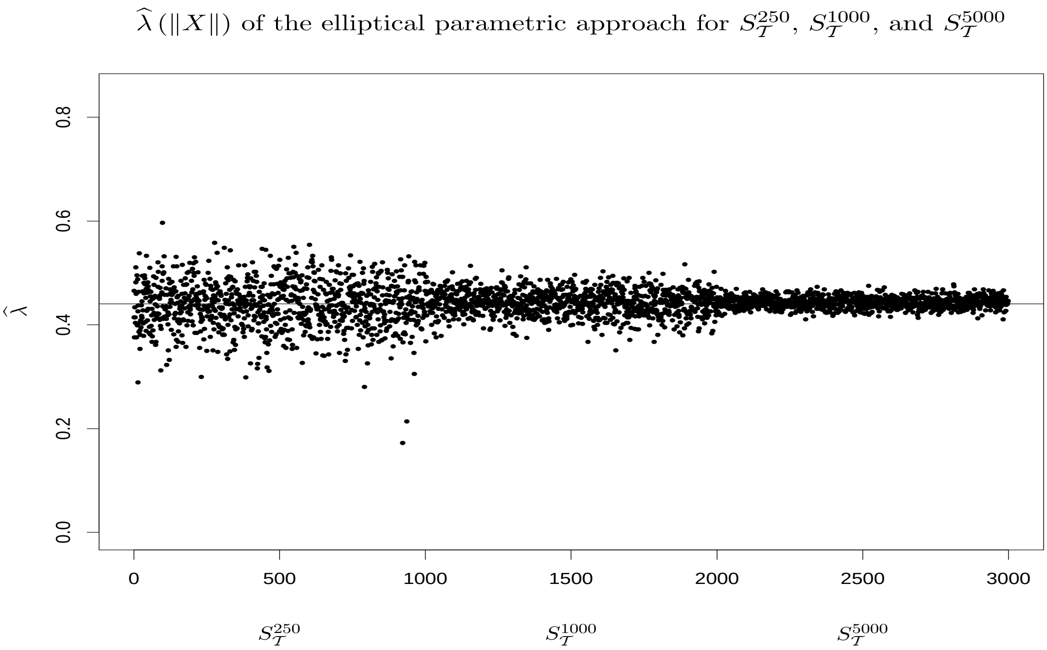


Figure 4.15: Consecutive estimates of λ from $S_{\mathcal{T}}^{250}$, $S_{\mathcal{T}}^{1000}$, and $S_{\mathcal{T}}^{5000}$ utilizing the estimator (4.39), where the tail index is estimated via $\|X\|_2$ and the correlation is estimated via Kendall's tau (the black line is the true value).

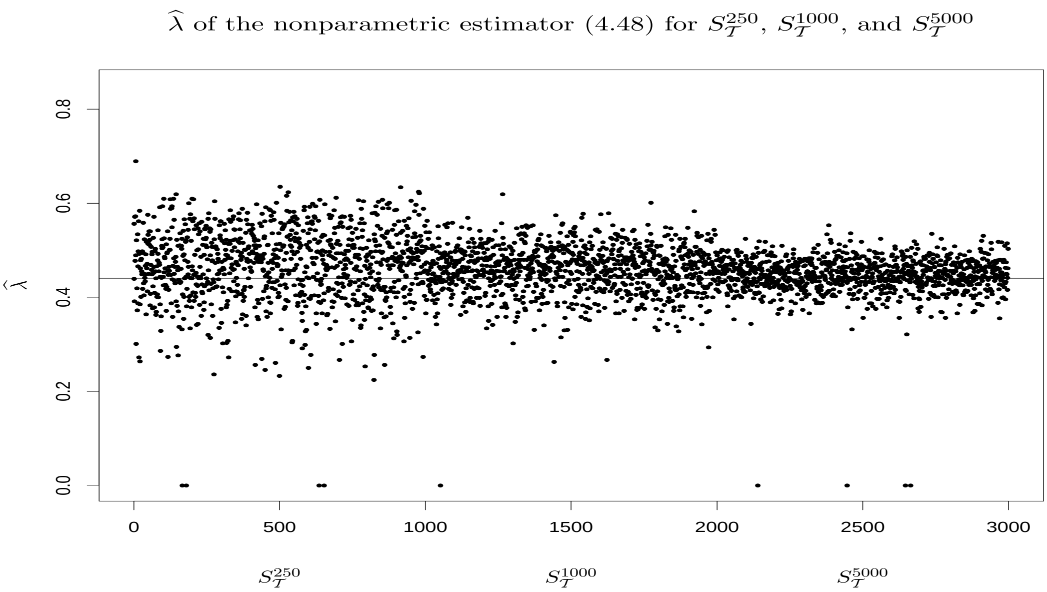


Figure 4.16: Consecutive estimates of λ from $S_{\mathcal{T}}^{250}$, $S_{\mathcal{T}}^{1000}$, and $S_{\mathcal{T}}^{5000}$ utilizing the nonparametric log-estimator (4.53) (the black line is the true value).

the condition $\sum_{i=k}^{k+l-1} |\lambda(k) - \lambda(i)| \leq 2\sigma(\Lambda)$. If such a plateau does not exist, the TDC is set to zero. Otherwise, if $\{\lambda(k^\star), \ldots, \lambda(k^\star + l - 1)\}$ denotes the respective plateau, then $\widehat{\lambda}_m := \frac{1}{l} \sum_{i=0}^{l-1} \lambda(k^\star + i)$.

4.11.3 Conclusions

The simulation and estimation results in Tables 4.6 and Table 4.7 reveal that in the case of high tail-dependence ($\lambda = 0.44$) all estimators under consideration, except the threshold method, yield a satisfying TDC estimation. As we expected, the finite sample bias of most estimations decreases with increasing sample size (exceptions are estimations from data-sets $\mathcal{F}$ and $\mathcal{G}$ via the semi-parametric method where we utilized empirical margins and the t-copula). The threshold method performs poorly as it usually overestimates the TDC which results in a large bias and therefore a large RMSE. The RMSE for all other estimation methods shows a better behavior with a small advantage towards the parametric elliptical methods. For the latter methods we observe quite small bias' for large sample sizes. However, in the case of tail independence, which is given by the data-set $\mathcal{H}$, the nonparametric estimators show a poor behavior. These estimators provide a large finite sample bias due to TDC overestimations which can be seen in a large RMSE and MESE. Obviously this overestimation is caused by the heuristic plateau-finding algorithm which is very sensitive towards high correlations. For lower correlation coefficient we usually do not observe this overestimation. Therefore a test on tail dependence prior to every estimation or an application of several estimators is recommended (cf. also Section 4.10.3 for other pitfalls).

For more empirical results regarding smaller values of the TDC we refer to Section 4.9.3.

Method	Data set	$\overline{\widehat{\lambda}_U}$	$\sigma\left(\widehat{\lambda}_U\right)$	$\text{RMSE}\left(\widehat{\lambda}_U\right)$	$\text{MESE}\left(\widehat{\lambda}_U\right)$
t-copula and	$S_{\mathcal{H}}^{250}$	0.1618	0.0817	0.1812	1.2184
empirical margins	$S_{\mathcal{H}}^{1000}$	0.1698	0.0413	0.1747	3.2315
cf. Section 4.10.1	$S_{\mathcal{H}}^{5000}$	0.1739	0.0187	0.1749	8.3676
and 4.11.2	$S_{\mathcal{T}}^{250}$	0.4281	0.0403	0.0422	0.0463
	$S_{\mathcal{T}}^{1000}$	0.4374	0.0204	0.0206	0.0098
	$S_{\mathcal{T}}^{5000}$	0.4400	0.0092	0.0092	0.0000
	$S_{\mathcal{F}}^{250}$	0.4873	0.0316	0.0564	0.7831
	$S_{\mathcal{F}}^{1000}$	0.4940	0.0150	0.0555	2.6862
	$S_{\mathcal{F}}^{5000}$	0.4946	0.0065	0.0544	7.3265
	$S_{\mathcal{G}}^{250}$	0.3905	0.0437	0.0664	0.5214
	$S_{\mathcal{G}}^{1000}$	0.3922	0.0212	0.0529	1.4914
	$S_{\mathcal{G}}^{5000}$	0.3919	0.0097	0.0497	4.1237
Elliptical approach	$S_{\mathcal{H}}^{250}$	0.2031	0.0588	0.2114	2.5959
via $\|X\|_2$	$S_{\mathcal{H}}^{1000}$	0.1815	0.0377	0.1854	3.9171
cf. Section 4.9.1	$S_{\mathcal{H}}^{5000}$	0.1575	0.0220	0.1590	6.2286
and 4.11.2	$S_{\mathcal{T}}^{250}$	0.4379	0.0465	0.0466	0.0012
	$S_{\mathcal{T}}^{1000}$	0.4432	0.0242	0.0243	0.0054
	$S_{\mathcal{T}}^{5000}$	0.4437	0.0109	0.0113	0.0398
	$S_{\mathcal{F}}^{250}$	0.4712	0.1228	0.1265	0.0302
	$S_{\mathcal{F}}^{1000}$	0.4122	0.0955	0.0996	0.0428
	$S_{\mathcal{F}}^{5000}$	0.4298	0.0450	0.0463	0.0277
Elliptical approach	$S_{\mathcal{H}}^{250}$	0.1883	0.0691	0.2006	1.9027
via R_n^*	$S_{\mathcal{H}}^{1000}$	0.1663	0.0427	0.1717	3.0209
cf. Section 4.9.1	$S_{\mathcal{H}}^{5000}$	0.1412	0.0239	0.1432	4.9920
and 4.11.2	$S_{\mathcal{T}}^{250}$	0.4377	0.0594	0.0595	0.0007
	$S_{\mathcal{T}}^{1000}$	0.4429	0.0308	0.0308	0.0022
	$S_{\mathcal{T}}^{5000}$	0.4441	0.0199	0.0202	0.0152
	$S_{\mathcal{F}}^{250}$	0.4553	0.1380	0.1387	0.0051
	$S_{\mathcal{F}}^{1000}$	0.4029	0.0960	0.1031	0.0739
	$S_{\mathcal{F}}^{5000}$	0.4403	0.0407	0.0407	0.0000

Table 4.6: Statistical figures for various parametric, semi-parametric and nonparametric estimation methods.

Method	Data set	$\overline{\widehat{\lambda}_U}$	$\sigma\left(\widehat{\lambda}_U\right)$	$\text{RMSE}\left(\widehat{\lambda}_U\right)$	$\text{MESE}\left(\widehat{\lambda}_U\right)$
Threshold method	$S_{\mathcal{H}}^{250}$	0.1772	0.1869	0.2575	0.3780
cf. Section 4.10.2	$S_{\mathcal{H}}^{1000}$	0.1236	0.1306	0.1798	0.3768
and 4.11.2	$S_{\mathcal{H}}^{5000}$	0.0780	0.0859	0.1160	0.3505
	$S_{\mathcal{T}}^{250}$	0.4882	0.2043	0.2097	0.0263
	$S_{\mathcal{T}}^{1000}$	0.5388	0.1195	0.1546	0.2940
	$S_{\mathcal{T}}^{5000}$	0.5538	0.0708	0.1335	0.8852
	$S_{\mathcal{F}}^{250}$	0.5308	0.1967	0.2163	0.0996
	$S_{\mathcal{F}}^{1000}$	0.5606	0.1136	0.1652	0.4541
	$S_{\mathcal{F}}^{5000}$	0.5671	0.0689	0.1440	1.0907
	$S_{\mathcal{G}}^{250}$	0.3980	0.2021	0.2064	0.0215
	$S_{\mathcal{G}}^{1000}$	0.4150	0.1560	0.1580	0.0130
	$S_{\mathcal{G}}^{5000}$	0.4320	0.0971	0.0974	0.0034
Nonpara. method	$S_{\mathcal{H}}^{250}$	0.3636	0.1016	0.3775	2.7158
emp. copula est.	$S_{\mathcal{H}}^{1000}$	0.3056	0.0717	0.3139	3.3779
cf. Section 4.10.2	$S_{\mathcal{H}}^{5000}$	0.2390	0.0932	0.2565	1.7525
and 4.11.2	$S_{\mathcal{T}}^{250}$	0.4681	0.0800	0.0845	0.0574
	$S_{\mathcal{T}}^{1000}$	0.4587	0.0513	0.0545	0.0606
	$S_{\mathcal{T}}^{5000}$	0.4463	0.0431	0.0435	0.0087
	$S_{\mathcal{F}}^{250}$	0.5326	0.0758	0.1192	0.5718
	$S_{\mathcal{F}}^{1000}$	0.5326	0.0554	0.1074	0.9386
	$S_{\mathcal{F}}^{5000}$	0.4880	0.0694	0.0841	0.2113
	$S_{\mathcal{G}}^{250}$	0.4841	0.0796	0.0907	0.1397
	$S_{\mathcal{G}}^{1000}$	0.4650	0.0482	0.0541	0.1208
	$S_{\mathcal{G}}^{5000}$	0.4453	0.0603	0.0605	0.0030
Nonpara. method	$S_{\mathcal{H}}^{250}$	0.3142	0.0889	0.3266	3.6728
log estimator	$S_{\mathcal{H}}^{1000}$	0.2904	0.0538	0.2953	4.4844
cf. Section 4.10.2	$S_{\mathcal{H}}^{5000}$	0.2567	0.0377	0.2595	5.8742
and 4.11.2	$S_{\mathcal{T}}^{250}$	0.3925	0.0775	0.0912	0.1766
	$S_{\mathcal{T}}^{1000}$	0.4133	0.0491	0.0562	0.1438
	$S_{\mathcal{T}}^{5000}$	0.4240	0.0352	0.0389	0.1048
	$S_{\mathcal{F}}^{250}$	0.4210	0.0904	0.0924	0.0228
	$S_{\mathcal{F}}^{1000}$	0.4933	0.0612	0.0807	0.3191
	$S_{\mathcal{F}}^{5000}$	0.4802	0.0406	0.0567	0.3967
	$S_{\mathcal{G}}^{250}$	0.3761	0.1173	0.1338	0.1405
	$S_{\mathcal{G}}^{1000}$	0.4098	0.0448	0.0544	0.2121
	$S_{\mathcal{G}}^{5000}$	0.4229	0.0233	0.0293	0.2550

Table 4.7: Statistical figures for various parametric, semi-parametric and nonparametric estimation methods.

Chapter 5

Integrative multivariate distribution models

Multivariate generalized hyperbolic distributions represent an attractive family of distributions for multivariate data modelling. However, in a limited data environment, robust and fast estimation procedures are rare. In the first part of this chapter we propose another class of multivariate distributions belonging to affine-linear transformed random vectors with independent and generalized hyperbolic margins. These distributions possess good estimation properties and have attractive dependence structures. In addition we introduce the estimation and simulation procedures. Further, advantages and disadvantages of both types of distributions are discussed and illustrated via a simulation study. In the second part of this chapter we concentrate on a family of semi-parametric multivariate distributions which primarily consists of a subclass of elliptically contoured distributions. Here the multivariate (symmetric) generalized hyperbolic distributions form an important special case. In particular the focus is on normal variance mixtures with self-decomposable mixing distributions. Finally, we fit the semi-parametric distributions to several financial time series and discuss the Value-at-Risk for linear asset portfolios.

5.1 Multivariate (affine) generalized hyperbolic distributions

Data modelling with generalized hyperbolic distributions has become quite popular in various areas. Originally, Barndorff-Nielsen (1977) utilized this class of distributions to model grain size distributions of wind-blown sand. In econometrical finance the latter family of distributions has been used for multidimensional asset-return modelling. In this context, the generalized hyperbolic distribution replaces the Gaussian distribution which cannot model the fat tails and the distributional skewness of real financial data. References are Eberlein and Keller (1995), Bingham and Kiesel (2001b), and Eberlein (2001).

Multivariate generalized hyperbolic distributions (in short: MGH distributions) were introduced by Barndorff-Nielsen (1978). These distributions have attractive analytical and statistical properties whereas robust and fast parameter estimations turn out to be difficult in higher dimensions. Further, the MGH distribution functions possess no parameter constellation for which they are the product of their marginal distribution functions. However, many applications require the multivariate distribution function to model both: Marginal dependence and independence. Because of these and other insufficiencies (see also Section 5.1.4) we introduce and explore a new class of multivariate distributions, the so called multivariate affine generalized hyperbolic distributions (in short: MAGH distributions). This class of distributions has an appealing stochastic representation and, in contrast to the MGH distributions, the estimation and simulation algorithms simplify. Moreover, our simulation study reveals that the goodness-of-fit of the MAGH distribution is comparable to that of the MGH distribution. The one-dimensional margins of an MAGH distribution are even more flexible due to more flexibility in the parameter choice.

We show that the dependence structure of an MAGH distribution is appropriate for data modelling. In particular, the correlation matrix as an important dependence measure is more intuitive and easier to handle for an MAGH distribution than for an MGH distribution. Further, zero correlation implies independent margins of the MAGH distribution whereas the margins of an MGH distribution do not have this property. Moreover, in contrast to the MGH distributions, the MAGH distributions can model tail dependence or dependencies of extreme events.

On the other hand, MGH distributions embrace a large subclass of distributions which belong to the family of elliptically or spherically contoured distributions. The family of elliptically contoured distributions inherits many useful properties from the multivariate normal distribution, like closure properties for linear regression and for passing to marginal distributions (see Section 3.2.1). Moreover, various tests for statistical inference under the assumption of a normal distribution are also applicable in the elliptical world (see Section 5.2.4). Therefore it also depends on the kind of application which distribution to put in favor.

Sections 5.1.1 and 5.1.2 introduce the MGH and MAGH distributions and characterize the subclass of elliptically contoured distributions. In Section 5.1.3 we investigate the dependence structure of both distributions by utilizing the theory of copulae. Section 5.1.4 discusses the advantages and disadvantages of the MGH and MAGH distributions. The corresponding estimation and generation algorithms are elaborated in Sections 5.1.5 and 5.1.6. Finally, in Section 5.1.7, we present a detailed simulation study to illustrate the elaborated results. For further details and results we refer to Hrycej, Schmidt, and Stützle (2002).

5.1.1 Multivariate generalized hyperbolic model (MGH)

Originally, a subclass of MGH distributions, namely the hyperbolic distributions, has been introduced via so-called variance-mean mixtures of inverse Gaussian distributions (see also Section 5.1.6). This subclass suffers from not having hyperbolic distributed margins, i.e., the subclass is not closed with respect to passing to marginal distributions. Therefore and because of other theoretical aspects, Barndorff-Nielsen (1977) extended the class to the family of MGH distributions. Many different parametrical representations of MGH density functions can be found in the literature; see Blæsild and Jensen (1981), for example. The following representation of an MGH density function is appropriate in our context.

Definition 5.1.1 (MGH distribution) *An n-dimensional random vector X is said to be* multivariate generalized hyperbolic *(MGH) distributed with location vector $\mu \in \mathbb{R}^n$ and scaling matrix $\Sigma \in \mathbb{R}^{n\times n}$ if it has the following stochastic representation:*

$$X \stackrel{d}{=} A'Y + \mu \tag{5.1}$$

for some lower triangular matrix $A' \in \mathbb{R}^{n\times n}$ such that $A'A = \Sigma$ is positive-definite and Y possesses the density function

$$f_Y(y) = c\frac{K_{\lambda-n/2}(\alpha\sqrt{1+y'y})}{(1+y'y)^{n/4-\lambda/2}}e^{\alpha\beta' y}, \quad y \in \mathbb{R}^n. \tag{5.2}$$

The normalizing constant c is given by

$$c = \frac{\alpha^{n/2}\,(1-\beta'\beta)^{\lambda/2}}{(2\pi)^{n/2}K_\lambda(\alpha\sqrt{1-\beta'\beta})}, \tag{5.3}$$

where K_ν denotes the modified Bessel-function of the third kind with index ν (cf. Magnus, Oberhettinger, and Soni (1966), pp. 65) and the parameter domain is $\|\beta\|_2 < 1$, $\alpha > 0$ and $\lambda \in \mathbb{R}$ ($\|\cdot\|_2$ denotes the Euclidian norm). The family of n-dimensional generalized hyperbolic distributions is denoted by $MGH_n(\mu, \Sigma, \omega)$, where $\omega := (\lambda, \alpha, \beta)$.

An important property of the above parameterization of the MGH density function is its invariance under affine-linear transformations. For $\lambda = (n+1)/2$ we obtain the *multivariate hyperbolic* density and for $\lambda = -1/2$ the *multivariate normal inverse Gaussian* density. Hence $\lambda = 1$ leads to hyperbolically distributed one-dimensional margins. It can be shown that MGH distributions with $\lambda = 1$ are closed with respect to passing to marginal distributions and to affine-linear transformations. The latter subclass turns out to be important for practical applications (see also Section 5.1.7 of the present thesis).

An MGH distribution belongs to the class of elliptically contoured distributions if and only if $\beta = (0, \dots, 0)'$. In this case the density function of X can be represented as

$$f_X(x) = |\Sigma|^{-1/2} g((x-\mu)'\Sigma^{-1}(x-\mu)), \quad x \in \mathbb{R}^n, \tag{5.4}$$

for some density generator function $g : \mathbb{R}_+ \to \mathbb{R}_+$. Consequently, the random vector Y in (5.1) is spherically distributed. The density generator g in (5.4) is given by $g(u) = cK_{\lambda-n/2}(\alpha\sqrt{1+u})/(1+u)^{n/4-\lambda/2}$, $u \in \mathbb{R}$, with some normalizing constant c. For a detailed treatment of elliptically contoured distributions, see Fang, Kotz, and Ng (1990a) or Cambanis, Huang, and Simons (1981) (see also Section 3.2).

Remark. Usually the following representation of an MGH density is found in the literature:

$$\bar{f}_X(x) = \bar{c}\frac{K_{\bar{\lambda}-n/2}(\bar{\alpha}\sqrt{\bar{\delta}^2 + (x-\bar{\mu})'\bar{\Sigma}^{-1}(x-\bar{\mu})})}{\left(\bar{\alpha}^{-1}\sqrt{\bar{\delta}^2 + (x-\bar{\mu})'\bar{\Sigma}^{-1}(x-\bar{\mu})}\right)^{n/2-\bar{\lambda}}} e^{\bar{\beta}'(x-\bar{\mu})}, \quad x \in I\!R^n, \tag{5.5}$$

with some normalizing constant $\bar{c}$. The domain of variation[1] of the parameter vector $\bar{\omega} = (\bar{\lambda}, \bar{\alpha}, \bar{\delta}, \bar{\beta})$ is as follows: $\bar{\lambda}, \bar{\alpha} \in I\!R$, $\bar{\beta}, \bar{\mu} \in I\!R^n$, $\bar{\delta} \in I\!R_+$, $\bar{\beta}'\bar{\Sigma}\bar{\beta} < \bar{\alpha}^2$ and $\bar{\Sigma} \in I\!R^{n\times n}$ being a positive-definite matrix with determinant $|\bar{\Sigma}| = 1$. The one-to-one mapping between the parameter vector ω corresponding to (5.2) and $\bar{\omega}$ corresponding to (5.5) is given by: $\lambda = \bar{\lambda}$, $\mu = \bar{\mu}$, $\alpha = \bar{\alpha}\bar{\delta}$, $\beta = 1/\bar{\alpha} \cdot \bar{A}\bar{\beta}$, $\bar{A}'\bar{A} = \bar{\Sigma}$, and $\Sigma = \bar{\delta}^2 \cdot \bar{\Sigma}$.

5.1.2 Multivariate affine generalized hyperbolic model (MAGH)

A disadvantage of multivariate generalized hyperbolic distributions (and of many other families of multivariate distributions) is that the margins X_i of $X = (X_1, \ldots, X_n)'$ are not independent for any choice of the scaling matrix Σ. In other words, they do not allow the modelling of phenomena where random variables result as the sum of independent random variables. This shortcoming is serious since the independence may be an undisputable property of the problem for which the stochastic model is sought. Furthermore, in case of asymmetry (i.e., $\beta \neq 0$) the covariance matrix is in a complex relationship with the matrix Σ which is shown in the next section.

Therefore we propose an alternative concept. Instead of a multivariate generalized hyperbolic distribution, a distribution is considered which is composed of n independent margins with univariate generalized hyperbolic distributions with zero location and unit scaling. Such a canonical random vector is then subject to an affine-linear transformation. As a consequence, the transformation matrix can be modelled proportionally to the square root of the covariance matrix inverse even in the asymmetric case. This property holds, for example, for multivariate normal distributions.

Definition 5.1.2 (MAGH distribution) *An n-dimensional random vector X is said to be* multivariate affine generalized hyperbolic *(MAGH) distributed with location vector $\mu \in I\!R^n$ and scaling matrix $\Sigma \in I\!R^{n\times n}$ if it has the following stochastic representation:*

$$X \stackrel{d}{=} A'Y + \mu \tag{5.6}$$

for some lower triangular matrix $A \in I\!R^{n\times n}$ such that $A'A = \Sigma$ is positive-definite and the random vector $Y = (Y_1, \ldots, Y_n)'$ consists of mutually independent random variables $Y_i \in MGH_1(0, 1, \omega_i)$, $i = 1, \ldots, n$. In particular the one-dimensional margins of Y are generalized hyperbolic distributed. The family of n-dimensional affine generalized hyperbolic distributions is denoted by $MAGH_n(\mu, \Sigma, \omega)$, where $\omega := (\omega_1, \ldots, \omega_n)$ and $\omega_i := (\lambda_i, \alpha_i, \beta_i)'$, $i = 1, \ldots, n$.

[1] This representation omits the limiting distributions obtained at the boundary of the parameter space; see e.g. Blæsild and Jensen (1981)

Observe that an MAGH distribution has independent margins in case the scaling matrix Σ equals the identity matrix I. However, no MAGH distribution belongs to the class of elliptically contoured distributions for dimension $n \geq 2$ which is illustrated by the density contour-plots in Figure 5.2.

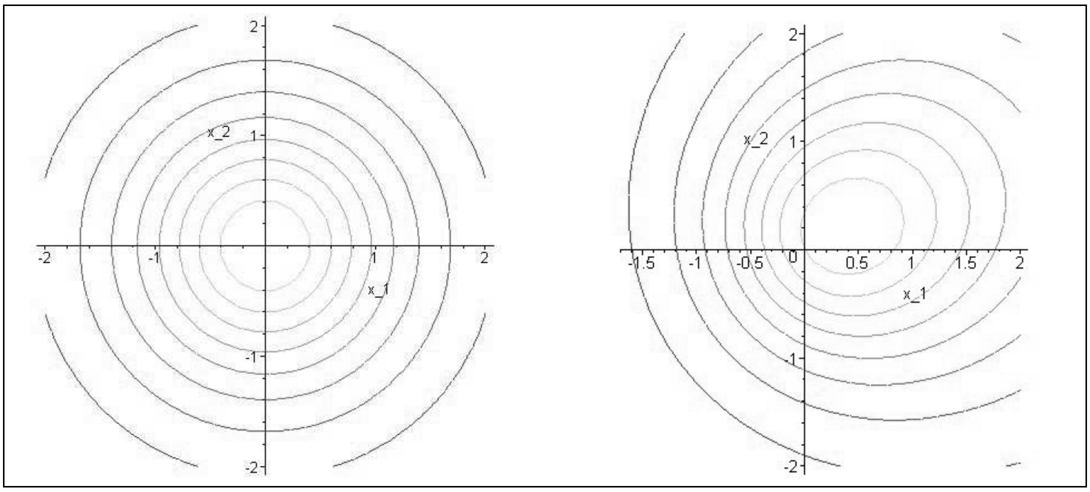


Figure 5.1: Contour-plots of the bivariate density function of an $MGH_2(0, I, \omega)$ distribution with parameters $\lambda = 1,\ \alpha = 1$ and $\beta = (0, 0)'$ (left figure), $\beta = (0.5, 0.25)'$ (right figure)

General affine transformations. The consideration of the lower triangular matrix A' in the stochastic representations (5.1) and (5.6) is essential, since any other decomposition of the scaling matrix Σ would lead to a different class of distributions. This phenomenon and a possible extension are discussed below.

Only the elliptically contoured subclass of the MGH distributions is invariant with respect to different decompositions $A'A = \Sigma$. In particular, all decompositions of the scaling matrix Σ lead to the same distribution since they always enter the characteristic function in the form $\Sigma = A'A$ (cf. Section 3.2.1). Equation (5.4) also justifies the latter property. However, in the asymmetric or general affine case this equivalence does not hold anymore. In this case, for example, the matrix A can be sought via a singular value decomposition

$$A = UWV' \tag{5.7}$$

where W is a diagonal matrix having the square roots of eigenvalues of $\Sigma = A'A$ on its diagonal and where the matrix V consists of the corresponding eigenvectors of Σ. The matrices W and V are directly determined from Σ whereas the matrix U might be some arbitrary matrix with orthonormal columns (rotation and flip). However, the most common case, of course, is $U = I$. Here the matrix A is directly computed from Σ utilizing its eigenvalues and eigenvectors. Consequently, every margin of Y is distributed according to a linear combination of the margins of X determined by the principal components (i.e., the eigenvectors) of the covariance matrix Σ:

$$Y = A'^{-1}X = W^{-1}V'X. \tag{5.8}$$

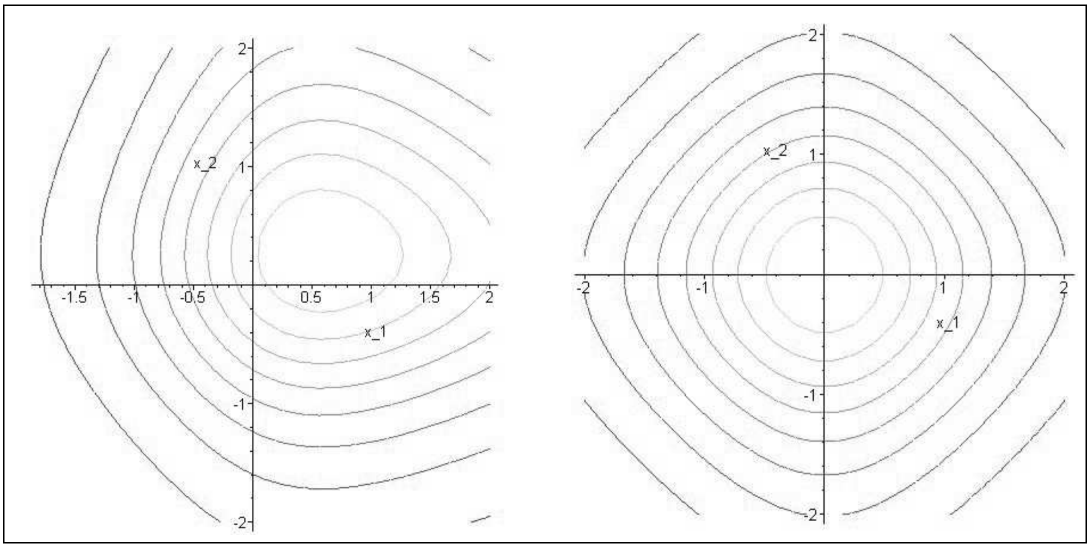


Figure 5.2: Contour-plots of the bivariate density function of an $MAGH_2(0, I, \omega)$ distribution with parameters $\lambda = (1,1)'$, $\alpha = (1,1)'$ and $\beta = (0,0)'$ (left figure), $\beta = (0.5, 0.25)'$ (right figure).

5.1.3 Dependence structures of MGH and MAGH distributions

In this section we investigate and compare the dependence structures of the MGH distribution and the MAGH distribution. We consider the corresponding copulae and various dependence measures, like the covariance and correlation coefficients, Kendall's tau and the tail-dependence coefficient.

The MGH and MAGH copulae. According to Definitions 5.1.1 and 5.1.2, the MGH and MAGH distributions are represented by affine-linear transformations of random vectors following a standardized MGH distribution and MAGH distribution, respectively. Leaving the affine-linear transformation aside, we are interested in the dependence structure (copula) of the underlying random vector. In particular we set the scaling matrix $\Sigma = I$ and $\mu = 0$. Note that the copula of an MGH or MAGH distribution does not even depend on the location vector, i.e., μ is not a copula parameter.

Theorem 5.1.3 *Let $Y \in MGH_n(0, I, \omega)$. Then the copula density function of Y is given by*

$$c(u_1, \ldots, u_n)$$
$$= c\frac{K_{\lambda-n/2}(\alpha\sqrt{1+y'y})}{(1+y'y)^{n/4-\lambda/2}} \prod_{i=1}^{n} \frac{(1+y_i^2)^{1/4-\lambda/2}}{K_{\lambda-1/2}\left(\alpha\sqrt{1+y_i^2}\right)} \frac{\exp(\alpha\beta' y)}{\exp(\prod_{i=1}^{n} \alpha\beta_i y_i)}\Bigg|_{y_i = F_i^{-1}(u_i)}$$

for $u_i \in [0,1]$, $i = 1, \ldots, n$, and some normalizing constant c. Here F_i refers to the distribution function of the one-dimensional margin Y_i, $i = 1, \ldots, n$.

Let $Y \in MAGH_n(0, I, \omega)$. Then the corresponding copula equals the independence copula, i.e, the copula density function is given by

$$c(u_1, \dots, u_n) \equiv 1, \quad u_i \in [0,1],\ i = 1, \dots, n. \tag{5.9}$$

Proof. Suppose $Y \in MGH_n(0, I, \omega)$. Then Y has a continuous and strictly positive density function f. Utilizing formula (2.11) we obtain the copula density function by

$$c(u_1, \dots, u_n) = \left. \frac{f(y_1, \dots, y_n)}{\prod_{i=1}^n f_i(y_i)} \right|_{y_i = F_i^{-1}(u_i)}. \tag{5.10}$$

Inserting the density function (5.2) into (5.10) yields the first assertion. Assume now that $Y \in MAGH_n(0, I, \omega)$. Then Y has independent margins if and only if Y possesses the independence copula $C(u_1, \dots, u_n) = u_1 \cdot \ldots \cdot u_n$ according to Theorem 2.10.14 in Nelsen (1999). □

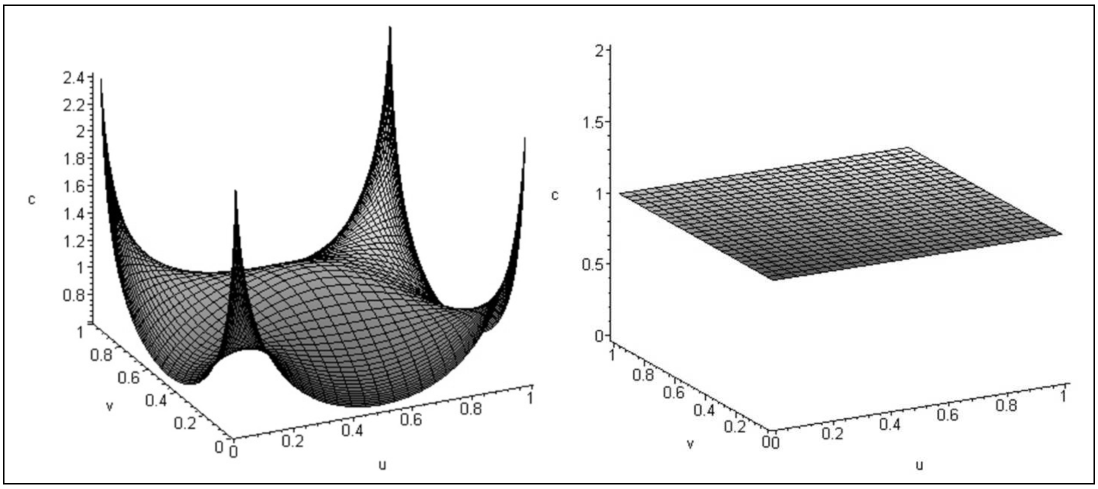


Figure 5.3: Copula-density function $c(u, v)$ of an $MGH_2(0, I, \omega)$ distribution (left figure) with parameters $\lambda = 1$, $\alpha = 1$ and $\beta = (0, 0)'$ and that of an $MAGH_2(0, I, \omega)$ distribution (right figure) with arbitrary parameter constellation.

Figure 5.3 clearly reveals that the independence copula of an MAGH distribution is quite different to the copula function of an MGH distribution. In particular, the MGH copula shows dependence in the limiting corners like in the far right-upper and lower-left quadrant. This property is investigated in more detail later, when we discuss the concept of tail dependence.

According to Theorem 2.10.12 in Nelsen (1999), any copula function is bounded from below (above) by the lower (upper) Fréchet bound, i.e.,

$$W(u_1, \dots, u_n) \le C(u_1, \dots, u_n) \le M(u_1, \dots, u_n), \tag{5.11}$$

where $W(u_1, \dots, u_n) = \max(u_1 + \ldots + u_n - n + 1, 0)$ and $M(u_1, \dots, u_n) = \min(u_1, \dots, u_n)$. Let $\Pi(u_1, \dots, u_n) = u_1 \cdot \ldots \cdot u_n$ be the product or independence copula. Note that M and Π are always copula functions. However, W is only a copula for dimension $n = 2$.

Theorem 5.1.4 (Limiting cases) *Let*

$$\Sigma^{(m)} := \begin{pmatrix} \sigma_{11}^{(m)} & \sigma_{12}^{(m)} \\ \sigma_{12}^{(m)} & \sigma_{22}^{(m)} \end{pmatrix} \in I\!R^{2\times 2},\ m \in I\!N,$$

be a sequence of symmetric positive-definite matrices and $\rho^{(m)} := \sigma_{12}^{(m)} / \sqrt{\sigma_{11}^{(m)}\sigma_{22}^{(m)}}$. *Suppose that* $X^{(m)} \in MGH_2(\mu, \Sigma^{(m)}, \omega)$ *or* $X^{(m)} \in MAGH_2(\mu, \Sigma^{(m)}, \omega)$ *for every* $m \in I\!N$, *and let* $C^{(m)}$ *denote the corresponding copula. Then*

i) $C^{(m)} \to M$ *pointwise if* $\rho^{(m)} \to 1,\ \sigma_{ij}^{(m)} \to \sigma_{ij} \neq 0,\ i,j = 1,2,$ *as* $m \to \infty$,

ii) $C^{(m)} \to W$ *pointwise if* $\rho^{(m)} \to -1,\ \sigma_{ij}^{(m)} \to \sigma_{ij} \neq 0,\ i,j = 1,2,$ *as* $m \to \infty$,

iii) if $X^{(m)} \in MGH_2(\mu, \Sigma^{(m)}, \omega)$, *then* $C^{(m)} \neq \Pi$ *for each parameter constellation, and*

iv) if $X^{(m)} \in MAGH_2(\mu, \Sigma^{(m)}, \omega)$, *then* $C^{(m)} = \Pi$ *if and only if* $\Sigma^{(m)} = I$.

Proof. i) For each $m \in I\!N$ the random vector $X^{(m)}$ possesses the stochastic representation $X^{(m)} \stackrel{d}{=} (A^{(m)})'Y + \mu$, with

$$(A^{(m)})' = \begin{pmatrix} \sqrt{\sigma_{11}^{(m)}} & 0 \\ \sigma_{12}^{(m)}/\sqrt{\sigma_{11}^{(m)}} & \sqrt{\sigma_{22}^{(m)}}\sqrt{1-(\rho^{(m)})^2} \end{pmatrix}$$

being the Cholesky decomposition of $\Sigma^{(m)}$. Note that the correlation coefficient of $X^{(m)} = (X_1^{(m)}, X_2^{(m)})'$ fulfills $\rho^{(m)} = \sigma_{12}^{(m)}/\sqrt{\sigma_{11}^{(m)}\sigma_{22}^{(m)}} \in (-1,1)$ for all $m \in I\!N$ since $\Sigma^{(m)}$ is positive-definite. If now $\rho^{(m)} \to 1,\ \sigma_{ij}^{(m)} \to \sigma_{ij} \neq 0,\ i,j = 1,2,$ as $m \to \infty$ then

$$X^{(m)} \stackrel{d}{\to} A'Y + \mu =: X \text{ as } m \to \infty \quad \text{with } A' = \begin{pmatrix} \sqrt{\sigma_{11}} & 0 \\ \sigma_{12}/\sqrt{\sigma_{11}} & 0 \end{pmatrix}.$$

Consequently for $X = (X_1, X_2)'$ and $\mu = (\mu_1, \mu_2)'$ we obtain $X_1 = (X_2 - \mu_2)\sigma_{11}/\sigma_{12} + \mu_1 = g(X_2)$ for some strictly increasing mapping $g : I\!R \to I\!R$ because $\sigma_{12} > 0$. Due to Theorem 2.4.3 in Nelsen (1999) the copula C of X is invariant under strictly increasing transformations of the margins because X possesses continuous marginal distribution functions. Therefore we conclude

$$C \equiv C_{X_1,X_2} \equiv C_{X_1,X_1} \equiv M.$$

Hence $C^{(m)} \to C \equiv M$ pointwise as $m \to \infty$ because $X^{(m)} \stackrel{d}{\to} X$ as $m \to \infty$ and the corresponding distribution functions are continuous.

ii) The second assertion can be shown analogously, using the fact that

$$C \equiv C_{X_1,X_2} \equiv C_{X_1,g(-X_1)} \equiv C_{X_1,-X_1} \equiv W$$

for some strictly increasing mapping $g : \mathbb{R} \to \mathbb{R}$.

iii) Suppose $X^{(m)} \in MGH_2(\mu, \Sigma^{(m)}, \omega)$. Then $X^{(m)}$ possesses the product copula if and only if it has independent margins (see Theorem 2.4.2 in Nelsen (1999)). According to Definition 5.1.1, $X^{(m)}$ does not have independent margins if $\Sigma^{(m)}$ is not a diagonal matrix. Thus it suffices to consider $\Sigma^{(m)} = I$. Further we can put $\beta = 0$ since β has no influence on the factorization of the density function of an MGH distribution (see formula (5.2)). However, in that case $X^{(m)}$ belongs to the family of elliptically contoured distributions. Therefore Theorem 4.11 in Fang, Kotz, and Ng (1990a) implies that $X^{(m)}$ possesses independent margins if and only if $X^{(m)}$ is has a bivariate normal distribution. Since normal distributions and MGH distributions are disjoint classes of distributions, the assertion follows.

iv) According to Definition 5.1.2, MAGH distributions have independent margins if and only if $\Sigma = I$. □

Remark. The results of Theorem 5.1.4 can be extended to n-dimensional MGH and MAGH distributions. However, for $n \geq 3$ the lower Fréchet bound is not a copula function anymore; see Theorem 2.10.13 in Nelsen (1999) for an interpretation of the lower Fréchet bound in that case.

The covariance and correlation matrix. Among the large number of dependence measures for multivariate random vectors the covariance and the correlation matrix are still the most favorite ones in most practical applications. Many multivariate models in finance are based on these linear dependence measures which are appropriate under the assumption of a normal distribution. However, as we have already mentioned, several empirical investigations reject the hypothesis of a multivariate normal distribution to be suited for financial data-modelling (see also Section 5.2). Consequently, new distribution models like the present MGH and MAGH model have been introduced where dependence measures describing only linear dependence between bivariate random vectors should be considered with care. Often, "scale-invariant" dependence measures like Kendall's tau seem to be more appropriate; for more details we refer to Embrechts, McNeil, and Straumann (1999). Recall, in Section 3.3.1 we showed that the correlation matrix is still a reasonable dependence measure in the elliptically contoured world.

Theorem 5.1.5 (Mean and covariance for MGH distributions)
Let $X \in MGH_n(\mu, \Sigma, \omega)$ and define $R_{\lambda,i}(x) := \frac{K_{\lambda+i}(x)}{x^i K_\lambda(x)}$. Then the mean vector and the covariance matrix of X are given by

$$E[X] = \mu + \alpha R_{\lambda,1}(\sqrt{\alpha^2(1-\beta'\beta)})\ A'\beta$$

and

$$\begin{aligned} Cov[X] &= R_{\lambda,1}(\sqrt{\alpha^2(1-\beta'\beta)})\,\Sigma \\ &+ \Big[R_{\lambda,2}(\sqrt{\alpha^2(1-\beta'\beta)}) - R^2_{\lambda,1}(\sqrt{\alpha^2(1-\beta'\beta)})\Big]\ \frac{A'\beta\beta' A}{1-\beta'\beta} \end{aligned}$$

For the symmetric case $\beta = (0,\dots,0)'$ and $\lambda = 1$, the mean vector and the covariance matrix of X simplify to $E[X] = 0$ and $Cov[X] = K_2(\alpha)/(\alpha K_1(\alpha))\cdot\Sigma$.

Proof. Let $X \in MGH_n(\bar\mu, \bar\Sigma, \bar\omega)$ with parameter representation as in (5.5). Then X is distributed like a variance-mean mixture of some multivariate normal distribution, i.e., $X|(Z=z) \sim N(\bar\mu + z\bar\Sigma\bar\beta, z\bar\Sigma)$, where the mixing random variable Z is distributed according to a generalized inverse Gaussian distribution $GIG(\bar\lambda, \bar\delta, \sqrt{\bar\delta^2(\bar\alpha^2 - \bar\beta'\bar\Sigma\bar\beta)})$ (see e.g. Barndorff-Nielsen, Kent, and Sørensen (1982b)). Hence,

$$E[X] = \bar\mu + \bar\Sigma\bar\beta E_{GIG}[Z]$$

and

$$E[XX'] = \bar\mu\bar\mu' + (\bar\Sigma + \bar\mu\bar\beta'\bar\Sigma + \bar\Sigma\bar\beta\bar\mu')E_{GIG}[Z] + \bar\Sigma\bar\beta\bar\beta'\bar\Sigma E_{GIG}[Z^2].$$

Therefore, the covariance matrix of X is given by

$$Cov[X] = \bar\Sigma E_{GIG}[Z] + \bar\Sigma\bar\beta\bar\beta'\bar\Sigma Var_{GIG}[Z].$$

According to Eberlein and Prause (1999), mean and variance of the mixing random variable $Z \sim GIG(\bar\lambda,\, \bar\delta,\, \sqrt{\bar\delta^2(\bar\alpha^2 - \bar\beta'\bar\Sigma\bar\beta)})$ are given by

$$E_{GIG}[Z] = R_{\bar\lambda,1}(\sqrt{\bar\delta^2(\bar\alpha^2 - \bar\beta'\bar\Sigma\bar\beta)})$$

and

$$Var_{GIG}[Z] = \bar\delta^4\cdot[R_{\bar\lambda,2}(\sqrt{\bar\delta^2(\bar\alpha^2 - \bar\beta'\bar\Sigma\bar\beta)}) - R^2_{\bar\lambda,1}(\sqrt{\bar\delta^2(\bar\alpha^2 - \bar\beta'\bar\Sigma\bar\beta)})].$$

Utilizing the parameter mapping $\bar\delta^2(\bar\alpha^2 - \bar\beta'\bar\Sigma\bar\beta) = \alpha^2(1-\beta'\beta)$, $\bar\Sigma\bar\beta\bar\delta^2 = \Sigma\alpha A^{-1}\beta$ and $\bar\delta^4\bar\Sigma\bar\beta\bar\beta'\bar\Sigma = A'\beta\beta' A$ yields the assertion. □

Theorem 5.1.6 (Mean and covariance for MAGH distributions)
Let $X \in MAGH_n(\mu, \Sigma, \omega)$. Then the mean vector and the covariance matrix are given by

$$E[X] = A'e_Y + \mu \quad \textit{and} \quad Cov[X] = A'CA,$$

respectively, where $e_Y = (E[Y_1],\dots,E[Y_n])'$ with $E[Y_i] = R_{\lambda_i,1}(\sqrt{\alpha_i^2(1-\beta_i^2)})\alpha_i\beta_i$ and $C = diag(c_{11},\dots,c_{nn})$ with

$$c_{ii} = R_{\lambda_i,1}(\sqrt{\alpha_i^2(1-\beta_i^2)}) + [R_{\lambda_i,2}(\sqrt{\alpha_i^2(1-\beta_i^2)}) - R^2_{\lambda_i,1}(\sqrt{\alpha_i^2(1-\beta_i^2)})]\frac{\beta_i^2}{1-\beta_i^2}.$$

The covariance matrix $Cov[X]$ is proportional to Σ if $\alpha = \alpha_i$, $\beta = \beta_i$ and $\lambda = \lambda_i$ for all $i = 1,\dots,n$.

Proof. The assertion follows immediately from Theorem 5.1.5 and equation (5.6). □

Kendall's tau. The correlation coefficient is a measure of linear dependence between two random variables and therefore it is not invariant under monotone increasing transformations. However, not only does "scale-invariance" present an undisputable requirement for a proper dependence measure in general (cf. Joe (1997), Chapter 5), but also in practice "scale-invariant" dependence measures play an increasing role in dependence modelling. Kendall's tau is the most famous one and therefore we determine it for MGH and MAGH distributions.

Definition 5.1.7 (Kendall's tau) *Let $X = (X_1, X_2)'$ and $\bar{X} = (\bar{X}_1, \bar{X}_2)'$ be independent bivariate random vectors with common continuous distribution function F and copula C. Kendall's tau is defined by*

$$\begin{aligned} \tau &= \mathbb{P}((X_1 - \bar{X}_1)(X_2 - \bar{X}_2) > 0) - \mathbb{P}((X_1 - \bar{X}_1)(X_2 - \bar{X}_2) < 0) \\ &= 4 \int_{[0,1]^2} C(u,v)\, dC(u,v) - 1. \end{aligned} \tag{5.12}$$

Theorem 5.1.8 *Let $\rho \in (-1, 1)$ be the correlation coefficient of X_1 and X_2.*
i) If $X \in MGH_2(\mu, \Sigma, \omega)$ with $\beta = 0$, then

$$\tau = \frac{2}{\pi} \arcsin(\rho). \tag{5.13}$$

ii) If $X \in MAGH_2(\mu, \Sigma, \omega)$ with stochastic representation $X \stackrel{d}{=} A'Y + \mu,\ A'A = \Sigma$, then for $\rho \neq 0$

$$\tau = \frac{4}{|c|} \int_{\mathbb{R}^2} f_{Y_1}(x_1) \left(\frac{x_2 - x_1}{c} \right) \cdot \int_{-\infty}^{x_1} F_{Y_2} \left(\frac{x_2 - z}{c} \right) f_{Y_1}(z)\, dz d(x_1, x_2) - 1, \tag{5.14}$$

where $c := sgn(\rho)\sqrt{1/\rho^2 - 1}$. Further $\tau = 0$ for $\rho = 0$.

Proof. i) Suppose $X \in MGH_2(\mu, \Sigma, \omega)$ with parameter $\beta = 0$. Then X belongs to the family of elliptically contoured distributions and the assertion follows by Theorem 2 in Lindskog, McNeil, and Schmock (2001).

ii) Suppose $X \in MAGH_2(\mu, \Sigma, \omega)$ with stochastic representation $X \stackrel{d}{=} A'Y + \mu,\ A'A = \Sigma$ and copula C. In particular

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \quad \text{and} \quad A' = \begin{pmatrix} \sqrt{\sigma_{11}} & 0 \\ \sigma_{12}/\sqrt{\sigma_{11}} & \sqrt{\sigma_{22}}\sqrt{1-\rho^2} \end{pmatrix}$$

with ρ being the correlation coefficient of X_1 and X_2. For $\rho = 0$, the assertion follows by Theorem 5.1.9 in Nelsen (1999) due to the independence of X_1 and X_2. For the remaining case we can assume $\rho > 0$ as for $\rho < 0$ the assertion is shown similarly. According to Theorem 5.1.3 in Nelsen (1999), Kendall's tau is a copula property what justifies to put $\mu = 0$. Further, Kendall's tau is invariant under strictly increasing transformations of the margins (see Theorem 5.1.9 in Nelsen (1999)) and therefore we may set

$$A' = \begin{pmatrix} 1 & 0 \\ 1 & c \end{pmatrix} \quad \text{with } c := \sqrt{1/\rho^2 - 1}.$$

Consequently, the distribution function of $X = (X_1, X_2)'$ has the form

$$F_X(x_1, x_2) = \mathbb{P}(Y_1 \leq x_1, Y_1 + cY_2 \leq x_2) = \int_{-\infty}^{x_1} F_{Y_2}\left(\frac{x_2 - z}{c}\right) f_{Y_1}(z)\, dz$$

and the corresponding density function is

$$f_X(x_1, x_2) = \frac{1}{|c|} f_{Y_2}\left(\frac{x_2 - x_1}{c}\right) f_{Y_1}(x_1).$$

Thus, using the fact that

$$\tau = 4 \int_{[0,1]^2} C(u,v)\, dC(u,v) - 1 = 4 \int_{\mathbb{R}^2} F_X(x_1, x_2) f_X(x_1, x_2)\, d(x_1, x_2) - 1,$$

formula (5.14) is shown. □

Remark. A closed form expression of Kendall's tau for MAGH distributions (given by (5.14)) cannot be expected. However, since the density functions of Y_1 and Y_2 are explicitly given, formula (5.14) yields a tractable numerical solution.

The tail-dependence coefficient. The main emphasis in this thesis is put on concepts and measures of extremal dependence like tail dependence and the tail-dependence coefficient, respectively (see Section 2.2.3 for definitions). Figure 5.3 reveals that bivariate standardized MGH distributions show more evidence of dependence in the upper-right and lower-left quadrant of its distribution function than MAGH distributions. However, the following theorem shows that MGH distributions are always tail independent whereas non-standardized MAGH distributions can even model tail dependence. For the sake of simplicity we restrict ourselves to the symmetric case $\beta = 0$.

Theorem 5.1.9 *Let $\rho \in (-1, 1)$ be the correlation coefficient of the bivariate distributions considered below. Suppose $\beta = 0$. Then*
i) the $MGH_2(\mu, \Sigma, \omega)$ distributions are upper and lower tail-independent,
ii) the $MAGH_2(\mu, \Sigma, \omega)$ distributions are upper and lower tail-independent if $\alpha_2 < \alpha_1\sqrt{1/\rho^2 - 1}$ or $\rho \leq 0$, and
iii) the $MAGH_2(\mu, \Sigma, \omega)$ distributions are upper and lower tail-dependent if $\alpha_2 > \alpha_1\sqrt{1/\rho^2 - 1}$ and $\rho > 0$.

In order to prove the theorem we first investigate the tail behavior of the univariate symmetric MGH and MAGH distributions. The tail of the distribution function F, as always, is denoted by $\overline{F} := 1 - F$.

Definition 5.1.10 (Semi-heavy tails) *A continuous (symmetric) function $g : \mathbb{R} \to (0, \infty)$ is called semi-heavy tailed (or exponentially tailed) if it satisfied*

$$g(x) \sim c|x|^\nu \exp(-\eta|x|) \quad \text{as } x \to \pm\infty \tag{5.15}$$

with $\nu \in \mathbb{R}$, $\eta > 0$ and some positive constant c. The class of (symmetric) semi-heavy tailed functions is denoted by $L_{\nu,\eta}$.

Lemma 5.1.11 *Let f be a density function such that $f \in L_{\nu,\eta}$, $\nu \in \mathbb{R}$, $\eta > 0$. Then the corresponding distribution function F possesses the same asymptotic behavior as its density, i.e., $F(x) \sim \bar{c}|x|^\nu \exp(-\eta|x|)$ as $x \to -\infty$ and $\bar{F}(x) \sim \bar{c}x^\nu \exp(-\eta x)$ as $x \to \infty$ for some positive constant $\bar{c}$; write $F \in L_{\nu,\eta}$.*

Proof. Consider e.g. the tail function $\bar{F}$. Applying partial integration we obtain

$$\begin{aligned} \bar{F}(x) &= \int_x^\infty f(u)\, du \sim c \int_x^\infty u^\nu \exp(-\eta u)\, du \\ &= c\eta x^\nu \exp(-\eta x) + \eta c\nu \int_x^\infty u^{\nu-1} \exp(-\eta u)\, du. \end{aligned}$$

Thus, the proof is complete if we show that

$$\int_x^\infty u^{\nu-1} \exp(-\eta u)\, du \Big/ x^\nu \exp(-\eta x) = o(1) \quad \text{as } x \to \infty.$$

Rewriting the latter quotient yields

$$0 \le \frac{1}{x} \int_x^\infty \left(\frac{u}{x}\right)^{\nu-1} \exp(-\eta(u-x))\, du = \frac{1}{x} \int_0^\infty \left(\frac{u+x}{x}\right)^{\nu-1} \exp(-\eta u)\, du.$$

The assertion is now immediate because

$$\left(\frac{u}{x}+1\right)^{\nu-1} \le (u+1)^{\nu-1} \text{ for } \nu \ge 1 \text{ and } \left(\frac{u}{x}+1\right)^{\nu-1} \le 1 \text{ for } \nu < 1$$

and the corresponding integrals exist. □

The next lemma is quite useful; it states that the tail of the convolution of two semi-heavy tailed distributions is determined by the heavier tail.

Lemma 5.1.12 *Let F_1 and F_2 be distribution functions with $F_1 \in L_{\nu_1,\eta_1}$ and $F_2 \in L_{\nu_2,\eta_2}$ where $0 < \eta_2 < \eta_1$, $\nu_1, \nu_2 \in \mathbb{R}$. Then $F_1 * F_2 \in L_{\nu_2,\eta_2}$ and, moreover,*

$$\lim_{t\to\infty} \overline{F_1 * F_2}(t) / \overline{F}_2(t) = m_1 := \int_{-\infty}^\infty e^{\eta_2 u} dF_1(u). \tag{5.16}$$

Proof. For some fixed $s > 1$, we have

$$\begin{aligned}\overline{F_1 * F_2}(t) &= \int_{-\infty}^{\infty} \overline{F}_2(t-u)dF_1(u) \\ &= \int_{-\infty}^{t/s} \overline{F}_2(t-u)dF_1(u) - \int_{-\infty}^{t-t/s} \overline{F}_2(u)dF_1(t-u) \\ &= \int_{-\infty}^{t/s} \overline{F}_2(t-u)dF_1(u) - \Big[F_1(t/s) - 1 - \int_{-\infty}^{t-t/s} F_2(u)dF_1(t-u)\Big] \\ &= \int_{-\infty}^{t/s} \overline{F}_2(t-u)dF_1(u) + \int_{-\infty}^{t-t/s} \overline{F}_1(t-u)dF_2(u) + \overline{F}_2(t-t/s)\overline{F}_1(t/s),\end{aligned}$$

where the last equality follows by partial integration. Thus, dominated convergence yields

$$\begin{aligned}\lim_{t\to\infty} \overline{F_1 * F_2}(t)/\overline{F}_2(t) &= \lim_{t\to\infty} \int_{-\infty}^{t/s} \overline{F}_2(t-u)/\overline{F}_2(t)dF_1(u) \\ &= \int_{-\infty}^{\infty} e^{\eta_2 u} dF_1(u) =: m_1 < \infty\end{aligned}$$

because

$$0 \le \int_{-\infty}^{t-t/s} \overline{F}_1(t-u)/\overline{F}_2(t)dF_2(u) \le \frac{\overline{F}_1(t/s)}{\overline{F}_2(t)} \cdot \overline{F}_2(t-t/s) \to 0 \quad \text{as } t \to \infty.$$

A consequence of the symmetric tails of F_1 and F_2 is that $\lim_{t\to-\infty} F_1 * F_2(t)/F_2(t) = m_1$. Hence, $F_1 * F_2 \in L_{\nu_2,\eta_2}$ is proven. □

According to Barndorff-Nielsen and Blæsild (1981), the univariate MGH distributions have semi-heavy tails, in particular

$$MGH_1(0,1,\omega) \sim c|x|^{\lambda-1} \exp((\mp\alpha + \alpha\beta)x) \text{ as } x \to \pm\infty \tag{5.17}$$

with some positive constant c. Hence, in the symmetric case $\beta = 0$ we obtain $MGH_1(0,1,\omega) \in L_{\nu,\eta}$ with $\nu = \lambda - 1$ and $\eta = \alpha$. Now we are ready to prove Theorem 5.1.9.

Proof of Theorem 5.1.9. We only show upper tail-dependence and upper tail-independence, respectively, as the lower pendant is obtained similarly. Recall that tail dependence is a copula property and therefore we may put $\mu = 0$.

i) Let $X \in MGH_2(0,\Sigma,\omega)$ with $\beta = 0$. In that case X belongs to the family of elliptically contoured distributions. According to Theorem 3.2.13, in Section 3.2.3 the assertion follows because of the exponentially-tailed density generator.

ii) Let $X \in MAGH_2(0,\Sigma,\omega)$ with stochastic representation $X \overset{d}{=} A'Y$ and Cholesky matrix $A' = \begin{pmatrix} a_{11} & 0 \\ a_{12} & a_{22} \end{pmatrix}$. Note that $a_{11}, a_{22} > 0$. Then

$$
\begin{aligned}
&\mathbb{P}(X_2 > F_{X_2}^{-1}(v) \mid X_1 > F_{X_1}^{-1}(v)) \\
&\quad = \frac{\mathbb{P}(a_{11}Y_1 > F_{X_1}^{-1}(v), a_{12}Y_1 + a_{22}Y_2 > F_{X_2}^{-1}(v))}{\mathbb{P}(a_{11}Y_1 > F_{X_1}^{-1}(v))} \\
&\quad = \frac{1}{1-v}\mathbb{P}(Y_1 > F_{Y_1}^{-1}(v), a_{12}Y_1 + a_{22}Y_2 > F_{X_2}^{-1}(v)) \\
&\quad = \frac{1}{1-v}\int_{-\infty}^{\infty} \mathbb{P}(y > F_{Y_1}^{-1}(v), a_{12}y + a_{22}Y_2 > F_{X_2}^{-1}(v))f_{Y_1}(y)\, dy \\
&\quad = \frac{1}{1-v}\int_{F_{Y_1}^{-1}(v)}^{\infty} \mathbb{P}(Y_2 > (F_{X_2}^{-1}(v) - a_{12}y)/a_{22})f_{Y_1}(y)\, dy =: I
\end{aligned}
$$

because Y_1 and Y_2 are independent random variables. If $\rho \leq 0$ then $a_{12} \leq 0$ and upper tail-independence immediately follows by dominated convergence.

Consider now $\rho > 0$ and therefore $a_{12} > 0$. Let $\alpha_2 < \alpha_1\sqrt{1/\rho^2 - 1}$. Then $\alpha_2/a_{22} < \alpha_1/a_{12}$. For all $\varepsilon \in (0, 1]$ we conclude with $u := 1 - v$ that

$$
\begin{aligned}
I &\leq \varepsilon + \frac{1}{u}\int_{F_{Y_1}^{-1}(1-u)}^{F_{Y_1}^{-1}(1-u\varepsilon)} \mathbb{P}(Y_2 > (F_{X_2}^{-1}(1-u) - a_{12}y)/a_{22})f_{Y_1}(y)\, dy \\
&\leq \varepsilon + (1-\varepsilon)\mathbb{P}(Y_2 > (F_{X_2}^{-1}(1-u) - a_{12}F_{Y_1}^{-1}(1-u\varepsilon))/a_{22}). \qquad (5.18)
\end{aligned}
$$

Due to (5.17) we know that $F_{a_{12}Y_1} \in L_{\nu_1,\eta_1}$ with $\nu_1 = \lambda_1 - 1$ and $\eta_1 = \alpha_1/a_{12}$. Thus, Lemma 5.1.12 gives $F_{X_2} \in L_{\nu_2,\eta_2}$ with $\nu_2 = \lambda_2 - 1$ and $\eta_2 = \alpha_2/a_{22}$ as $0 < \eta_2 < \eta_1$. Then the probability in (5.18) converges to zero as $u \to 0^+$ if $F_{X_2}^{-1}(1-u) - a_{12}F_{Y_1}^{-1}(1-u\varepsilon) = F_{a_{12}Y_1+a_{22}Y_2}^{-1}(1-u) - F_{a_{12}Y_1}^{-1}(1-u\varepsilon) \to \infty$ as $u \to 0^+$. Put $x_u := F_{a_{12}Y_1}^{-1}(1-u\varepsilon)$ and $y_u := F_{a_{12}Y_1+a_{22}Y_2}^{-1}(1-u)$. Then

$$
u = \frac{1}{\varepsilon}\overline{F}_{a_{12}Y_1}(x_u) = \overline{F}_{X_2}(y_u) \sim \frac{c_1}{\varepsilon}x_u^{\nu_1}\exp(-\eta_1 x_u) \sim c_2 y_u^{\nu_2}\exp(-\eta_2 y_u) \qquad (5.19)
$$

as $u \to 0^+$ and therefore $x_u, y_u \to \infty$. The asymptotic behavior (5.19) implies $y_u - x_u \to \infty$ as $u \to 0^+$ because $0 < \eta_2 < \eta_1$. Hence, upper tail-independence is shown.

iii) Now suppose $\rho > 0$ and $\alpha_2 > \alpha_1\sqrt{1/\rho^2 - 1}$ which yields $a_{12} > 0$ and $\alpha_2/a_{22} > \alpha_1/a_{12}$. According to Lemma 5.1.12 we have $F_{a_{12}Y_1} \in L_{\nu_1,\eta_1}$ and $F_{X_2} \in L_{\nu_1,\eta_1}$ with $\nu_1 = \lambda_1 - 1$ and $\eta_1 = \alpha_1/a_{12}$. Notice that with $u := 1 - v$

$$
I \geq \mathbb{P}(Y_2 > (F_{X_2}^{-1}(1-u) - a_{12}F_{Y_1}^{-1}(1-u\varepsilon))/a_{22}) \qquad (5.20)
$$

Further

$$
u = \overline{F}_{a_{12}Y_1}(x_u) = \overline{F}_{X_2}(y_u) \sim c_1 x_u^{\nu_1}\exp(-\eta_1 x_u) \sim c_2 y_u^{\nu_1}\exp(-\eta_1 y_u).
$$

Hence

$$\frac{c_2}{c_1}\Big(\frac{y_u}{x_u}\Big)^{\nu_1}\exp(-\eta_1(y_u-x_u)) \to 1 \text{ as } u \to 0^+.$$

Suppose that $\limsup_{u\to 0^+}(y_u - x_u) = \infty$. If $\nu_1 \ge 0$, then the fact

$$\begin{aligned}\liminf_{u\to 0^+} \frac{c_2}{c_1}\Big(\frac{y_u}{x_u}\Big)^{\nu_1}&\exp(-\eta_1(y_u-x_u)) \\ &\le \liminf_{u\to 0^+}\frac{c_2}{c_1}(2(y_u-x_u))^{\nu_1}\exp\Big(-\eta_1(y_u-x_u)\Big) = 0 \text{ as } u\to 0^+\end{aligned}$$

would lead to a contradiction. On the other hand, if $\nu_1 < 0$ then

$$\liminf_{u\to 0^+} \frac{c_2}{c_1}\Big(\frac{y_u}{x_u}\Big)^{\nu_1}\exp(-\eta_1(y_u-x_u)) \le \liminf_{u\to 0^+}\frac{c_2}{c_1}\exp\Big(-\eta_1(y_u-x_u)\Big) = 0 \text{ as } u\to 0^+$$

would also lead to a contradiction. Therefore we conclude that $\liminf_{u\to 0^+}\mathbb{P}(Y_2 > (F_{X_2}^{-1}(1-u) - a_{12}F_{Y_1}^{-1}(1-u\varepsilon))/a_{22}) > 0$ as Y_2 is supported on $\mathbb{R}$. Finally, the limit $\lim_{v\to 1-}\mathbb{P}(X_2 > F_{X_2}^{-1}(v) \mid X_1 > F_{X_1}^{-1}(v))$ exists due to the convexity of $\overline{F}_{X_1}$ and $\overline{F}_{X_2}$ for large arguments. □

Remark. Additionally to Theorem 5.1.9 it can be shown that an $MGH_2(\mu,\Sigma,\omega)$ distribution (with $\beta = 0$) is tail independent if $\alpha_2 = \alpha_1\sqrt{1/\rho^2 - 1}$ and $\lambda_2 < \lambda_1$.

More details regarding modelling, properties, and estimation of tail dependence are provided in Chapters 3 and 4.

5.1.4 MGH versus MAGH: Advantages and disadvantages

In this section we list and compare some advantages and disadvantages of MGH distributions and MAGH distributions. We start with the *distributional flexibility* to fit real data. An outstanding property of MAGH distributions is that, after an affine-linear transformation, all one-dimensional margins can be fitted independently via flexible generalized hyperbolic distributions. In contrast to this, the one-dimensional margins of MGH distributions are not that adaptable since the parameters α and λ relate to the entire multivariate distribution and determine the strong structural behavior (see Definition 5.1.1). However, this structure causes a large subclass of MGH distributions to belong to the family of elliptically contoured distributions which inherit many useful statistical and analytical properties from normal distributions.

Regarding the *dependence structure*, the MAGH distributions may have independent margins for any parameter constellations $\omega = (\omega_1, \ldots, \omega_n)$ (see Theorem 5.1.4). In particular, they also support models which are based on a linear combination of independent factors. In contrast, the MGH distributions are not capable of modelling independent margins. They even yield "extremal" dependencies for bivariate distributions having correlation zero. Moreover, the correlation matrix of MAGH distributions is proportional to the scaling matrix Σ within a large subclass of asymmetric MAGH distributions (see Theorem 5.1.6). Whereas Σ is hardly to interpret for MGH distributions. Moreover, the copula of MAGH distributions, being the dependence structure of an affine-linear transformed random vector with independent components, is quite illustrative and possesses many appealing modelling properties. On the other hand, the copula structure of MGH distributions suffers from inflexibility due to the strong symmetry. Regarding the tail-dependence property, the MAGH distributions can model tail dependence whereas MGH distributions are always tail independent. Therefore we propose the MAGH distributions within the field of risk management.

Sections 5.1.5 and 5.1.7 reveal that in contrast to MGH distributions, *parameter estimation* for MAGH distributions is considerably simpler and more robust. Even in an asymmetric environment, the parameters of MAGH distributions can be identified in a two stage procedure which has a considerable computational advantage in higher dimensions. The same procedures can be applied for elliptically contoured MGH distributions ($\beta = 0$). The *simulation* algorithms for MGH and MAGH distributions turn out to be equally efficient and fast, irrespective of the dimension.

Simulation studies show that both distributions fit the simulated and real data quite well. Thus, summarizing the above advantages and disadvantages, the MAGH distributions have much to recommend them regarding their parameter estimation, dependence structure, and random vector generation. However, it depends also on the kind of application and the user's taste which model to prefer.

5.1.5 Parameter estimation

MGH distributions

Minimizing the cross entropy. A general method to measure the similarity between two distribution or density functions f^* and f, respectively, is given by the *Kullback entropy* (see Kullback (1959)) which is defined as

$$H_K(f, f^*) := \int_{\mathbb{R}^n} f^*(x) \log \frac{f^*(x)}{f(x)} dx = \int_{\mathbb{R}^n} f^*(x) \log f^*(x) dx - \int_{\mathbb{R}^n} f^*(x) \log f(x) dx. \tag{5.21}$$

The Kullback entropy is always nonnegative and is zero only if the densities f and f^* are identical. In our context, this relationship is useful to measure the similarity between the "true" density f^* and its approximation f. Therefore, minimizing the Kullback entropy by varying f is one way to find a good approximation of the "true" density f^*. Note that the first term in (5.21) is constant and can be dropped. This leads to the so-called *cross entropy* given by

$$H(f, f^*) := - \int_{\mathbb{R}^n} f^*(x) \log f(x) dx. \tag{5.22}$$

Obviously the integral in (5.22) cannot be evaluated without knowledge of f^*. Hence, the corresponding distribution is approximated by its empirical counterpart, i.e., by the empirical distribution function

$$F_e^*(x) = \frac{1}{m} \sum_{k=1}^{m} \mathbf{1}_{\{X_k \le x\}} \tag{5.23}$$

is considered, where $X_k,\ k = 1,\dots,m$, denotes a random sample with distribution function F^*. It is well known that this procedure leads to an unbiased estimator of $F^*(x)$ (see, for example, Parzen (1962)).

An approximation of the integral in (5.22) is now given by

$$H(f, f^*) \approx -\int_{I\!R^n} \log f(x) dF_e^*(x) = -\frac{1}{m}\sum_{k=1}^{m} \log f(X_k). \tag{5.24}$$

Thus, minimizing the cross entropy is approximately equivalent to maximizing the log-likelihood function.

Because of numerical reasons it can be advantageous to work with the standardized vector $y = B(x-\mu)$ where $B := A'^{-1}$. It is well known (see e.g. Stuart and Ord (1994)) that the density function f_X of $X \stackrel{d}{=} A'Y + \mu \in MGH(\mu, \Sigma, \omega)$ is given by

$$f_X(x) = f_{A'Y+\mu}(x) = |B| f_Y(y), \tag{5.25}$$

with $|B| > 0$ being the determinant of B. The following expression indicates the parameterization of the negative log-likelihood function L while utilizing the above standardization:

$$L(X, \eta) = -\sum_{k=1}^{m} \log(|B|) - \sum_{k=1}^{m} \log(f_Y(\omega, B(X_k - \mu))) \tag{5.26}$$

with parameters $\eta = (\mu, B, \omega)$ and $\omega = (\lambda, \alpha, \beta)$.

While the location vector μ can take arbitrary values in $I\!R^n$, B and ω are subject to various constraints. The matrix B must be triangular with positive diagonal such that $A'A$ is positive-definite. The parameter α is supposed to be positive and the vector β must fulfill $||\beta||_2 \leq 1$.

Many approaches are possible regarding the latter optimization problem, inter alia we mention two methods:

- constrained nonlinear optimization methods, and
- unconstrained nonlinear optimization methods after suitable parameter transformations.

We prefer the unconstrained approach due to robustness and efficiency reasons. The following parameter transformations are appropriate.

The matrix B can be sought of the form $B = UD$ where D is some diagonal matrix having strictly positive elements and U is a triangular matrix having only ones on its diagonal. In order to enforce the strict positivity of the diagonal elements $d_{ii}, i = 1,\dots,n$, of D, the following transformations are applied:

$$\begin{aligned} b_{ii} &= d_{ii} = e^{\nu_i},\ i = 1,\dots,n, \\ b_{ij} &= d_{ii}u_{ij} = e^{\nu_i}u_{ij}, i = 1,\dots,n-1;\ j = 2,\dots,n;\ j > i \end{aligned}$$

with unknown parameters ν_i and u_{ij}. The parameter α is estimated via the same exponential map. For the vector β we utilize the smooth transformation

$$\beta = \gamma \frac{1}{1 + \exp(-\|\gamma\|_2)} \cdot \frac{1}{\|\gamma\|_2}. \tag{5.27}$$

With the latter transformations we are now confronted with the new optimization problem

$$\min_{\bar{\eta} \in \Omega} L(X, \bar{\eta})$$

where $\bar{\eta}$ denote the transformed unconstrained parameters and $\Omega = I\!R^{(3+(n-1)/2)n+2}$. Since in general the objective function is non-convex, a global numerical optimization method is required. Optimization methods can be subdivided into deterministic and probabilistic methods. As far as deterministic methods are concerned, certain assumptions on the objective function have to be imposed (like Lipschitz conditions) in order to guarantee a termination of the optimization routine in finitely many steps. Probabilistic methods, in contrast, have better convergence properties but only with a certain likelihood. Here the assumptions on the objective function are rather weak.

The algorithm we use belongs to the probabilistic ones and consists of two characteristic phases:

1. In a *global phase* the objective function is evaluated at a random number of points being part of the search space.
2. In a *local phase* the samples are transformed to become candidates for local optimization routines.

Loosely speaking, the global phase is responsible for the convergence result of the optimization routine while the local phase determines the efficiency of the optimization. In the ideal (but hardly attainable) case, only a few local optimizations are necessary as there are attractors in the range of the objective function (e.g. regions where the minimum is reached by gradient descent). We mentioned already that the probabilistic algorithms find a global solution only with a given probability. Rinnooy Kan and Timmer (1987) show the convergence of probabilistic algorithms to the global minimum under certain conditions. This motivates a usage of the Multi-Level Single-Linkage procedure (see Rinnooy Kan and Timmer (1987)). Here the global optimization method generates random samples from the search space by identifying the samples belonging to the same objective function attractor and eliminating multiple ones. For the remaining samples in the local phase a conjugate gradient method (see Fletcher (1987)) is started. Further, a Bayesian stopping rule is applied in order to assess the probability that all attractors have been explored. For an account on the Bayesian stopping rule we refer the reader to Boender and Rinnooy Kan (1987).

MAGH distributions

The parameters of the MAGH distribution function can be easily identified in a two-stage procedure comprising the following steps:

1. Computing the sample covariance matrix S of the random vector X. Transforming X to the vector $Y = BX$ with independent margins. The matrix B is received via Cholesky decomposition $S^{-1} = B'B$.

2. Identifying the parameters λ_i, α_i and β_i which belong to the univariate marginal distributions of Y. The location vector μ can be received via $B^{-1}e$ with e_i being the location parameter of Y_i (see Theorem 5.1.5). The scaling matrix Σ equals ${B^{-1}}'DB^{-1}$ where the diagonal matrix D is determined by the scaling parameters of Y_i on its diagonal.

The latter procedure simplifies the complexity of the numerical optimization a lot.

Note that the parameters of MAGH distributions can be estimated analogously to the procedure described for MGH distributions. However, instead of the multivariate density (5.2), a product of n univariate densities

$$f_1(y) = c\frac{K_{\lambda-1/2}(\alpha\sqrt{1+y^2})}{(1+y^2)^{1/4-\lambda/2}}e^{\alpha\beta y} \tag{5.28}$$

with normalizing constant

$$c = \frac{\alpha^{1/2}\left(1-\beta^2\right)^{\lambda/2}}{(2\pi)^{1/2}K_\lambda(\alpha\sqrt{1-\beta^2})} \tag{5.29}$$

is considered. It is important for applications that the univariate densities are not necessarily identically parameterized. This means that the margins may have different parameters $\lambda_i, \alpha_i, \beta_i,\ i = 1, \ldots, n$. In other words, there is a considerable freedom of choosing the parameters λ, α and β. In addition to a parameterization similar to that for MGH distributions (i.e., the same λ and α for all one-dimensional margins and different β_i) two further extreme alternatives are possible:

- Minimum parameterization: Equal parameters λ, α and β for all one-dimensional margins.

- Maximum parameterization: Individual parameters λ_i, α_i and β_i for all one-dimensional margins.

The appropriate parameterization depends on the kind of application and the available data volume.

The optimization procedure presented in this section is a special case of an identification-algorithm for conditional distributions explored in Stützle and Hrycej (2001, 2002a, 2002b).

5.1.6 Sampling from MGH and MAGH distributions

Additionally to the estimation procedures described in Section 5.1.5, in this section we provide an efficient and self-contained random vector generation for the families of MGH and MAGH distributions. Fast sampling from an $MGH_n(\mu, \Sigma, \omega)$ distribution is possible via the following variance-mean mixture representation.

Let the random variable Z be distributed according to a generalized inverse Gaussian distribution with parameters λ, χ and ψ. In particular the latter family is referred to as the $GIG(\lambda, \chi, \psi)$ distributions. Then, $X \in MGH_n(\mu, \Sigma, \omega)$ is conditionally normal distributed with mixing random variable Z, i.e., $X|(Z = z) \sim N_n(\mu + z\tilde{\beta}, z\Delta)$, where $\Delta \in \mathbb{R}^{n\times n}$ is a symmetric positive-definite matrix with determinant $|\Delta| = 1$ and $\mu, \tilde{\beta} \in \mathbb{R}^n$. The parameters Σ and $\omega = (\lambda, \alpha, \beta)$ are given by

$$\alpha = \sqrt{(\psi + \tilde{\beta}'\Delta\tilde{\beta})\chi},\ \beta = 1/\sqrt{\psi + \tilde{\beta}'\Delta\tilde{\beta}}L\tilde{\beta},\ \Sigma = \chi \cdot \Delta$$

with Cholesky decomposition $L'L = \Delta$. The inverse map is given by

$$\chi = |\Sigma|^{1/n},\ \psi = \alpha^2/|\Sigma|^{1/n} \cdot (1 - \beta'\beta),\ \Delta = \Sigma/|\Sigma|^{1/n},\ \text{and } \tilde{\beta} = \alpha \cdot (A)^{-1}\beta$$

with Cholesky decomposition $A'A = \Sigma$.

The sampling algorithm is now of the following form: A pseudo random number is sampled from a random variable Z having a generalized inverse Gaussian distribution with parameters λ, χ, and ψ. Then an n-dimensional random vector X being conditionally normal distributed with mean vector $\mu + Z\Delta\tilde{\beta}$ ("drift") and covariance matrix $Z\Delta$ (determinant $|\Delta| = 1$) is generated.

The density function of the generalized inverse Gaussian distribution $GIG(\lambda, \chi, \psi)$, is given by

$$f_Z(x) = c \cdot x^{\lambda-1} \exp\Big(-\frac{\chi}{2x} - \frac{\psi x}{2}\Big), \quad x > 0, \tag{5.30}$$

with normalizing constant $c = (\psi/\chi)^{\lambda/2}/(2K_\lambda(\sqrt{\psi\chi}))$. The range of the parameters is given by

i) $\chi > 0,\ \psi \geq 0$ if $\lambda < 0$, or

ii) $\chi > 0,\ \psi > 0$ if $\lambda = 0$, or

iii) $\chi \geq 0,\ \psi > 0$ if $\lambda > 0$.

The following efficient algorithm is formulated for multivariate generalized hyperbolic distributions $MGH_n(\mu, \Sigma, \omega)$ with parameter $\lambda = 1$. Section 5.1.7 justifies the restriction to that class of distributions. In this context the $GIG(\lambda, \chi, \psi)$ distribution is referred to as inverse Gaussian distribution. However, the algorithm can be easily extended to general λ. An empirical study shows that the algorithm outperforms the efficiency of the sampling algorithm proposed by Atkinson (1982) which suites to a larger class of distributions (see also Prause (1999), Section 4.6). Moreover, the algorithm avoids tedious minimization routines and time-consuming evaluations of the Bessel function K_λ. The generation utilizes a rejection method (see Ross (1997), pp. 565) with a three part rejection-envelop. We define the envelop $d : I\!R_+ \to I\!R_+$ by

$$d(x) := \begin{cases} d_1(x) = ca_1 \exp(b_1 x), & \text{if } 0 < x < x_1, \\ d_2(x) = ca_2, & \text{if } x_1 \le x < x_2, \\ d_3(x) = ca_3 \exp(-b_3 x), & \text{if } x_2 \le x < \infty, \end{cases} \tag{5.31}$$

with $a_i > 0,\ i = 1, \ldots, 3,\ b_i > 0,\ i = 1, 3$, and $x_1 \le x_2 \le x_3$ to be defined later. Let $z_i,\ i = 1, 3$ denote the inflection points and $z_2 = \sqrt{\chi/\psi}$ denote the mode of the unimodal density f_Z. Further we require

$$d_1(z_1) = f_Z(z_1), \quad d_2(z_2) = f_Z(z_2), \quad d_3(z_3) = f_Z(z_3). \tag{5.32}$$

The points $x_1 > 0$ and $x_2 > 0$ correspond to the intersection points of d_1, d_2 and d_2, d_3, respectively, i.e.,

$$d_1(x_1) = d_2(x_1), \qquad d_2(x_2) = d_3(x_2). \tag{5.33}$$

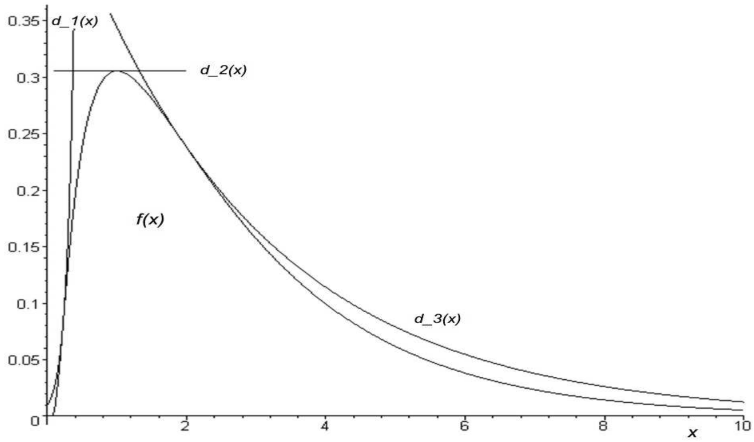


Figure 5.4: Three part envelop d for the inverse Gaussian density function f_Z with parameters $\chi = 1$, $\psi = 1$.

Primarily, the rejection method requires the generation of random numbers with density $s \cdot d(x)$ where the scaling factor s has to be computed in order to obtain a density function $s \cdot d(x), x > 0$. This scaling factor is derived below.

Pseudo algorithm for generating an inverse Gaussian random number:

1. Compute the zeros z_1, z_2 for $\psi^2 z^4 - 2\chi\psi z^2 - 4\chi z + \chi^2 = 0$.
2. Set $b_1 = (\chi/z_1^2 - \psi)/2$ and $a_1 = \exp(-\chi/z_1)$.
3. Set $a_2 = \exp(-\sqrt{\chi\psi})$.

4. If $(\psi - \chi/z_2^2)/2 > 0$ then
   Set $b_3 = (\psi - \chi/z_2^2)/2$ and $a_3 = \exp(-\chi/z_2)$
   Else Set $b_3 = \psi/2$ and $a_3 = 1$.

5. Set $x_1 = \ln(a_2/a_1)/b_1$ and $x_2 = -\ln(a_2/a_3)/b_3$.

6. Set
$$s = \Big(\frac{a_1}{b_1}\exp(b_1 x_1) - \frac{a_1}{b_1} + (x_2 - x_1)a_2 + \frac{a_3}{b_3}\exp(-b_3 x_2)\Big).$$

7. Set
$$k_1 = \frac{1}{s}\Big(\frac{a_1}{b_1}\exp(b_1 x_1) - \frac{a_1}{b_1}\Big) \quad \text{and} \quad k_2 = k_1 + \frac{1}{s}(x_2 - x_1)a_2.$$

8. Generate independent and uniformly distributed random numbers U and V on the interval $[0, 1]$.

9. If $U \le k_1$ goto step 10.
   ElseIf $k_1 < U \le k_2$ goto step 11.
   Else goto step 12.

10. Set
$$x = \frac{1}{b_1}\ln(\frac{b_1}{a_1}sU + 1).$$
    If
$$V \le \frac{f_Z(x)}{d_1(x)} = \frac{1}{a_1}\exp\Big(- (\frac{\chi x^{-1} + \psi x}{2} + b_1 x)\Big)$$
    Then Return x
    Else goto step 8.

11. Set
$$x = \frac{sU}{a_2} - \frac{a_1}{b_1 a_2}\Big(\exp(b_1 x_1) - 1\Big) + x_1.$$
    If
$$V \le \frac{f_Z(x)}{d_2(x)} = \frac{1}{a_2}\exp\Big(- (\frac{\chi x^{-1} + \psi x}{2})\Big)$$
    Then Return x
    Else goto step 8.

12. Set
$$x = -\frac{1}{b_3}\ln\Big[- \frac{b_3}{a_3}\Big\{sU - \frac{a_1}{b_1}\big(e^{b_1 x_1} - 1\big) - (x_2 - x_1)a_2 - \frac{a_3}{b_3}e^{-b_3 x_2}\Big\}\Big].$$
    If
$$V \le \frac{f_Z(x)}{d_3(x)} = \frac{1}{a_3}\exp\Big(- (\frac{\chi x^{-1} + \psi x}{2}) + b_3 x\Big)$$
    Then Return x
    Else goto step 8.

Remark. In order to generate a sequence of inverse Gaussian random numbers repeat step 8.

So far we have generated random numbers from an univariate inverse Gaussian distribution. We turn now to the generation of multivariate generalized hyperbolic random vectors. For this we exploit the above introduced mixture representation.

Pseudo algorithm for generating an MGH vector:

1. Set $\Delta = L'L$ via Cholesky decomposition.
2. Generate an inverse Gaussian random number Z with parameters χ and ψ.
3. Generate a standard normal random vector N.
4. Return $X = \mu + Z\Delta\tilde{\beta} + \sqrt{Z}L'N$.

Pseudo algorithm for generating an MAGH vector:

1. Set $\Sigma = A'A$ via Cholesky decomposition.
2. Generate a random vector Y with independent $MGH_1(0, 1, \omega_i)$, $i = 1, \ldots, n$, distributed components (see above).
3. Return $X = \mu + A'Y$.

Table 5.1 presents values of empirical efficiency of the MGH random vector generator for various parameter constellations. In our framework, efficiency is defined by the following ratio

$$\text{Efficiency} = \frac{\sharp \text{ of generated samples}}{\sharp \text{ of algorithm-passes including rejections}}\,.$$

χ/ψ	0.1	0.5	1	2	5	10
0.1	0.94	0.913	0.904	0.886	0.877	0.872
0.5	0.916	0.884	0.877	0.865	0.859	0.852
1	0.901	0.877	0.867	0.863	0.857	0.851
2	0.889	0.866	0.857	0.856	0.856	0.845
5	0.876	0.858	0.854	0.852	0.847	0.849
10	0.866	0.86	0.851	0.853	0.842	0.847

Table 5.1: Empirical efficiency of the MGH random number generator for $\lambda = 1$ and 10,000 generated samples.

5.1.7 Simulation and empirical study

A series of computational experiments with simulated data is performed in this section. The experiments disclose that

- MGH density functions with arbitrary parameter λ seem to have very close counterparts in the MGH-subclass with $\lambda = 1$,
- The MAGH model can closely approximate the MGH model with similar parameter values.

Arbitrary MGH distributions versus multivariate generalized hyperbolic distributions with $\lambda = 1$ (MH)

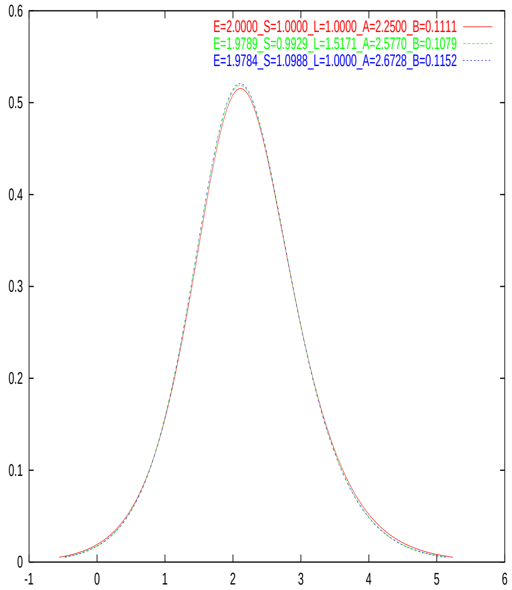
Figure 5.5 shows the identification results of four univariate distributions with random samples drawn from the following reference distributions. The identification takes place either under the assumption of an arbitrary generalized hyperbolic distribution or of a generalized hyperbolic distribution with fixed $\lambda = 1$ (in short: MH distribution). In both pictures on the left half of the figure, samples were drawn from the generalized hyperbolic distribution with parameter $\lambda = 1$. They illustrate the phenomenon that although the identification procedure for MGH distributions frequently produces $\lambda \neq 1$, the approximation of the density function remains good. The pictures on the right side illustrate the opposite case, namely, a good approximation of the MGH density function ($\lambda \neq 1$) via the MH density function ($\lambda = 1$).

Consider now an $MGH_2(\mu, \Sigma, \omega)$ distribution with $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$. The mutual tradeoff between λ, α and the scaling parameters $S_1 := \sqrt{\sigma_{11}}$ and $S_2 := \sqrt{\sigma_{22}}$ for the latter distribution function is shown in Table 5.2. While all samples were drawn from an MH distribution, MGH identifications usually lead to an overestimate of λ which is traded off by lower values of α, S_1 and S_2. In contrast to that, the parameter identifications for MH distributions are close to the reference values. However, the differences between the cross entropies are hardly discernible, showing that both parameter combinations correspond to densities which are close to each other.

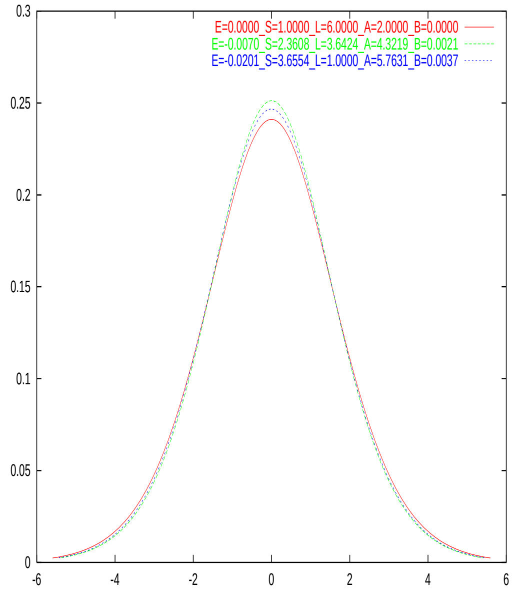
The following conclusions can be drawn:

- The fit of both distributions, MGH and MH distribution, measured by cross entropy and visual closeness of the plots, is satisfying. This implies the existence of multiple parameter constellations for MGH distribution functions which lead to quite similar density functions. Similar results have been observed for the corresponding tail functions.
- Generalized hyperbolic densities seem to have very close counterparts in the class of MH distributions (even for large λ). Therefore, the class of MH distributions will be sufficiently rich for our considerations.

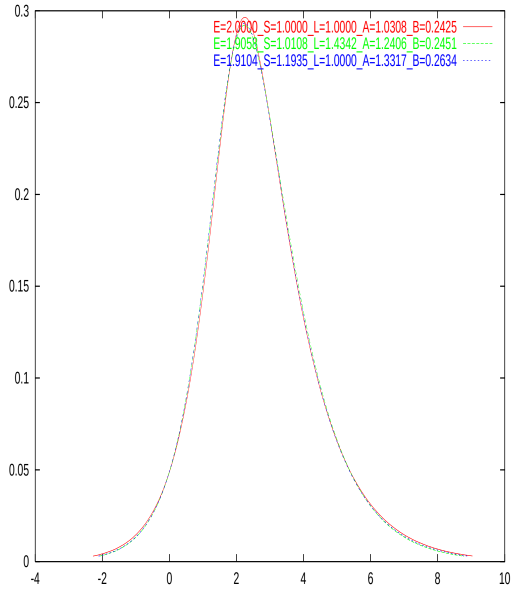
In view of the above results, only MH distributions and the corresponding MAH distributions (multivariate affine generalized hyperbolic distributions with $\lambda = 1$) will be considered in the next section.



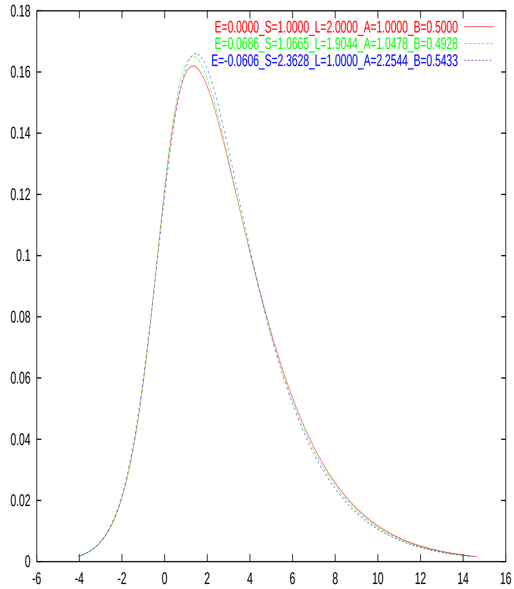
$(\mu, \sqrt{\Sigma}, \lambda, \alpha, \beta)' =$
$(2.00, 1.00, 1.00, 2.25, 0.11)'$ (solid line)
$(1.98, 0.99, 1.51, 2.58, 0.11)'$ (bright dotted line)
$(1.98, 1.10, 1.00, 2.67, 0.12)'$ (dark dotted line)



$(\mu, \sqrt{\Sigma}, \lambda, \alpha, \beta)' =$
$(0.00, 1.00, 6.00, 2.00, 0.00)'$ (solid line)
$(0.01, 2.36, 3.64, 4.32, 0.00)'$ (bright dotted line)
$(0.02, 3.66, 1.00, 5.76, 0.00)'$ (dark dotted line)



$(\mu, \sqrt{\Sigma}, \lambda, \alpha, \beta)' =$
$(2.00, 1.00, 1.00, 1.03, 0.24)'$ (solid line)
$(1.91, 1.01, 1.43, 1.24, 0.24)'$ (bright dotted line)
$(1.91, 1.19, 1.00, 1.33, 0.26)'$ (dark dotted line)



$(\mu, \sqrt{\Sigma}, \lambda, \alpha, \beta)' =$
$(0.00, 1.00, 2.00, 1.00, 0.50)'$ (solid line)
$(0.07, 1.07, 1.90, 1.05, 0.49)'$ (bright dotted line)
$(0.06, 2.36, 1.00, 2.25, 0.54)'$ (dark dotted line)

Figure 5.5: Univariate MGH and MH distributions. Reference densities (solid lines) and identified MGH and MH densities (dotted lines) averaged from 100 random samples for various parameter constellations .

MH distributions versus MAH distributions

The estimation of parameters is compared for the following three classes of bivariate distributions:

1. MH distributions.
2. MAH distributions with minimal parameter configuration (same value of α and β for each margin).
3. MAH distributions with maximal parameter configuration (different values of α_i and β_i for each margin).

All types of distributions in Table 5.3 have been identified from data which were sampled from an MH distribution. The two stage algorithm introduced in Section 5.1.5 has been used for the identification of the MAHmax model (determining first the sample correlation matrix, then transforming the variables, and finally identifying the univariate distributions). The identification results are provided in Table 5.3.

The following conclusions can be drawn:

- For all three models, most parameters show an acceptable fit regarding the sample bias and the sample standard deviation. The relative variability of the estimates increases with decreasing α (fatter tailed distributions). Such fatter tailed distributions seem to be more ill-posed with respect to the estimation of individual parameters.
- The differences between the parameter estimates obtained either for the MH, the MAHmin, or the MAHmax distribution are negligible (although the data are drawn from an MH distribution).
- The fit in terms of the cross entropy does not differ significantly between the various models. As expected, the MAHmax estimates are closer to the MH reference distribution than the MAHmin estimates are in terms of the cross entropy (Note that in one case they are even better than the MH estimates). The fitting capability of the MAHmax model comes at the expense of a larger variability and sometimes larger bias ("overlearning effect").

5.1.8 Application to financial data

The MGH and MAGH distributions have been fitted to various asset-return data. In particular, the following distributions have been used:

1. MGH/MH,
2. MAGH/MAH with minimum parameterization (denoted by (min)), that is, with all margins equally parameterized,

Table 5.2: Bivariate MGH and MH distribution. For each parameter λ, α, $S_1 = \sqrt{\sigma_{11}}$, and $S_2 = \sqrt{\sigma_{22}}$ the table lists the reference value, the sample mean $m(\cdot)$, and the sample standard deviation $\sigma(\cdot)$ of various parameter estimations derived from 100 samples of sample-size 1000 each. In the last column, the sample mean and the sample standard deviation of the corresponding cross entropy H are provided.

	λ	$m(\hat{\lambda})$	$\sigma(\hat{\lambda})$	α	$m(\hat{\alpha})$	$\sigma(\hat{\alpha})$	S_1	$m(\hat{S}_1)$	$\sigma(\hat{S}_1)$	S_2	$m(\hat{S}_2)$	$\sigma(\hat{S}_2)$	$m(\hat{H})$	$\sigma(\hat{H})$
MGH	1.0000	1.1061	0.0571	0.3200	0.2724	0.0867	0.3200	0.2574	0.0735	0.3200	0.2595	0.0746	3.4525	0.0438
MH	1.0000	1.0000	0.0000	0.3200	0.3286	0.1228	0.3200	0.3229	0.1078	0.3200	0.3243	0.1119	3.4177	0.3450
MGH	1.0000	1.3745	0.1651	1.1456	0.8758	0.2085	1.0000	0.7096	0.1341	1.0000	0.7141	0.1343	3.8683	0.0455
MH	1.0000	1.0000	0.0000	1.1456	1.1609	0.2922	1.0000	0.9954	0.1912	1.0000	1.0037	0.1931	3.8683	0.0454
MGH	1.0000	2.1779	0.5777	2.2400	1.8217	0.5151	1.0000	0.6897	0.1603	1.0000	0.6931	0.1621	2.5389	0.0411
MH	1.0000	1.0000	0.0000	2.2400	2.4543	0.6451	1.0000	1.0475	0.1859	1.0000	1.0522	0.1862	2.5390	0.0411

Table 5.3: Bivariate MH and MAH distribution. Reference value, sample bias and sample standard deviation (in brackets) are provided for each parameter estimate derived from 100 samples of sample-size 1000 each. Denote Dep.Par.:= $\sigma_{12}/(S_1 S_2)$. The last column lists the sample mean and the sample standard deviation of the cross entropy H.

Distribution	μ_1	μ_2	S_1	S_2	α_1	α_2	β_1	β_2	Dep.Par.	Cross Ent. H
	0.000 value:	0.000 value:	1.000 value:	1.000 value:	1.000 value:	1.000 value:	0.000 value:	0.000 value:	0.000 value:	
MAHmin	0.004 (0.085)	0.002 (0.087)	0.019 (0.208)	0.025 (0.213)	0.040 (0.265)	0.040 (0.265)	0.002 (0.032)	0.002 (0.032)	0.005 (0.037)	3.769 (0.039)
MAHmax	0.005 (0.111)	0.014 (0.112)	0.034 (0.323)	0.060 (0.300)	0.064 (0.398)	0.102 (0.385)	0.006 (0.045)	0.003 (0.044)	0.005 (0.048)	3.768 (0.039)
MH	0.003 (0.095)	0.007 (0.094)	0.023 (0.163)	0.029 (0.168)	0.042 (0.210)	0.042 (0.210)	0.005 (0.040)	0.001 (0.039)	0.003 (0.036)	3.753 (0.038)
	0.000 value:	0.000 value:	0.320 value:	0.320 value:	0.320 value:	0.320 value:	0.000 value:	0.000 value:	0.000 value:	
MAHmin	0.022 (0.830)	0.051 (0.644)	0.001 (0.126)	0.003 (0.131)	0.008 (0.137)	0.008 (0.137)	0.004 (0.115)	0.004 (0.115)	0.019 (0.109)	3.457 (0.112)
MAHmax	0.060 (0.464)	0.015 (0.200)	0.000 (0.205)	0.031 (0.194)	0.011 (0.214)	0.017 (0.203)	0.020 (0.111)	0.008 (0.122)	0.004 (0.214)	3.382 (0.142)
MH	0.071 (0.697)	0.100 (1.015)	0.003 (0.108)	0.004 (0.112)	0.009 (0.123)	0.009 (0.123)	0.004 (0.035)	0.007 (0.067)	0.012 (0.105)	3.418 (0.345)
	0.000 value:	0.000 value:	1.155 value:	1.155 value:	2.236 value:	2.236 value:	0.000 value:	0.000 value:	0.500 value:	
MAHmin	0.001 (0.120)	0.001 (0.075)	0.139 (0.238)	0.143 (0.191)	0.077 (0.626)	0.077 (0.626)	0.003 (0.042)	0.003 (0.042)	0.055 (0.027)	2.687 (0.040)
MAHmax	0.002 (0.110)	0.002 (0.127)	0.085 (0.294)	0.056 (0.320)	0.318 (1.083)	0.270 (0.969)	0.004 (0.064)	0.001 (0.053)	0.004 (0.145)	2.686 (0.040)
MH	0.003 (0.118)	0.002 (0.083)	0.159 (0.224)	0.128 (0.177)	0.128 (0.603)	0.128 (0.603)	0.003 (0.058)	0.000 (0.048)	0.054 (0.026)	2.680 (0.040)
	2.000 value:	4.000 value:	1.155 value:	1.155 value:	1.258 value:	1.258 value:	0.229 value:	0.512 value:	0.500 value:	
MAHmin	0.147 (0.201)	0.759 (0.129)	0.098 (0.253)	0.006 (0.234)	0.224 (0.279)	0.224 (0.279)	0.145 (0.034)	0.138 (0.034)	0.050 (0.029)	4.210 (0.050)
MAHmax	1.470 (0.173)	0.498 (0.171)	0.504 (0.511)	0.521 (0.192)	0.134 (0.483)	0.400 (0.303)	0.154 (0.039)	0.050 (0.014)	0.144 (0.123)	4.175 (0.048)
MH	0.201 (0.117)	0.260 (0.082)	0.031 (0.214)	0.186 (0.174)	0.182 (0.240)	0.182 (0.240)	0.128 (0.038)	0.042 (0.013)	0.007 (0.027)	4.128 (0.045)

3. MAGH/MAH with maximum parameterization (denoted by (max)), that is, with each margin individually parameterized.

For some of these distributions we have also estimated the symmetric (denoted by (sym)) pendant, i.e., $\beta = 0$. Further, we provide estimations following the affine transformation method discussed at the end of Section 5.1.2 (denoted by (PC)).

The results are presented in Table 5.4. The dependence parameter Dep.Par. refers to the sub-diagonal elements of the normed matrix Σ, i.e., Dep.Par.:= $\sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}} = \sigma_{ij}/(S_i S_j)$.

The maximum parameterized MAGH distribution frequently reach the best fit in terms of the cross entropy. Considerable variations between λ and α among the various models can be observed. This may lead to a similar variation of the scaling parameters which frequently behave contrary to the variation of the shape parameters.

Due to the total or approximate symmetry of some distribution models, the dependence parameters can be roughly interpreted as "correlation coefficients" for MGH and MAGHmin distributions. They are even close to the corresponding sample correlation coefficient (column "Corr."). Note that such an interpretation is not possible for the MAGHmax model, due to the different parameterization of the one-dimensional margins.

Summarizing the results we have found an appealing class of multidimensional distributions for modelling of high-dimensional data, namely the multivariate affine generalized hyperbolic distributions. These distributions are attractive regarding the estimation of unknown parameters and the random vector generation. Further we showed that this class possesses an appealing dependence structure and we proved several related dependence properties. Finally, an extensive simulation study showed the flexibility and robustness of the introduced model. Thus, the usage of multivariate affine generalized hyperbolic distributions is favorable for the dependence and data modelling within many statistical and financial applications.

Table 5.4: Estimations from two-dimensional asset-return data (DAX-CAC, DAX-DOW, NIKKEI-CAC).

Data	Model	Location μ	Scaling S	Dep.Par.	Corr.	λ	α	β	CrossEnt
DaxCac	MAGHmin	0.0012 0.0008	0.0062 0.0060	0.689	0.709	0.889	0.675	-0.034	-6.183
DaxCac	MAHmin	0.0011 0.0008	0.0062 0.0059	0.688	0.709	1.000	0.699	-0.029	-6.183
DaxCac	MAGHmin sym.	0.0004 0.0003	0.0056 0.0054	0.689	0.709	1.002	0.631	0.000	-6.182
DaxCac	MAGHmaxPC	0.0012 0.0005	0.0103 0.0046	0.941	0.709	0.812 0.732	0.192 1.363	0.037 0.000	-6.208
DaxCac	MAHmaxPC	0.0015 0.0007	0.0095 0.0042	0.898	0.709	1.000 1.000	0.261 1.315	0.039 0.000	-6.208
DaxCac	MAGHmaxPC sym.	0.0002 0.0001	0.0107 0.0048	0.999	0.709	0.635 0.645	0.050 1.427	0.000 0.000	-6.275
DaxCac	MAGHmax	0.0010 0.0001	0.0086 0.0082	0.957	0.709	0.873 0.800	0.247 1.490	-0.039 0.035	-6.201
DaxCac	MAHmax	-0.9333 -0.6267	0.0071 0.0071	0.999	0.709	1.000 1.000	0.050 1.302	0.600 0.037	-6.389
DaxCac	MAGHmax sym.	0.0001 0.0000	0.0085 0.0085	1.000	0.709	0.635 0.639	0.050 1.463	0.000 0.000	-6.265
DaxCac	MGH	0.0010 0.0005	0.0045 0.0044	0.672	0.709	1.134	0.535	-0.038 -0.017	-6.213
DaxCac	MH	0.0011 0.0006	0.0059 0.0058	0.673	0.709	1.000	0.689	-0.044 -0.019	-6.213
DaxCac	MGH sym.	0.0005 0.0003	0.0043 0.0042	0.672	0.709	1.143	0.504	0.000 0.000	-6.212
DaxDow	MAGHmin	0.0025 0.0013	0.0137 0.0097	0.505	0.498	0.891	1.285	-0.067	-5.743
DaxDow	MAHmin	0.0021 0.0011	0.0124 0.0088	0.503	0.498	1.000	1.169	-0.056	-5.743
DaxDow	MAGHmin sym.	0.0002 0.0001	0.0137 0.0097	0.500	0.498	0.759	1.208	0.000	-5.741
DaxDow	MAGHmaxPC	0.0009 0.0006	0.0145 0.0092	0.287	0.498	0.694 0.654	1.014 1.623	-0.001 0.000	-5.746
DaxDow	MAHmaxPC	0.0012 0.0002	0.0122 0.0082	0.373	0.498	1.000 1.000	0.868 1.561	0.032 0.000	-5.746
DaxDow	MAGHmaxPC sym.	0.0000 0.0001	0.0147 0.0093	0.277	0.498	0.650 0.657	1.013 1.634	0.000 0.000	-5.746
DaxDow	MAGHmax	0.0016 0.0010	0.0137 0.0114	0.477	0.498	1.160 0.751	1.194 1.825	-0.060 -0.014	-5.732
DaxDow	MAHmax	0.0017 0.0010	0.0146 0.0107	0.420	0.498	1.000 1.000	1.276 1.786	-0.063 -0.013	-5.732
DaxDow	MAGHmax sym.	0.0000 0.0002	0.0172 0.0119	0.397	0.498	0.650 0.653	1.423 1.878	0.000 0.000	-5.732
DaxDow	MGH	0.0020 0.0003	0.0135 0.0097	0.489	0.498	1.087	1.359	-0.077 -0.010	-5.757
DaxDow	MH	0.0021 0.0004	0.0137 0.0098	0.489	0.498	1.000	1.346	-0.079 -0.015	-5.757
DaxDow	MGH sym.	0.0001 0.0001	0.0130 0.0093	0.488	0.498	1.097	1.289	0.000 0.000	-5.755
NikkeiCac	MAGHmin	0.0000 0.0004	0.0028 0.0027	0.196	0.236	0.870	0.272	-0.011	-5.884
NikkeiCac	MAHmin	-0.0000 0.0004	0.0024 0.0023	0.199	0.236	1.000	0.245	-0.008	-5.884
NikkeiCac	MAGHmin sym.	-0.0001 0.0003	0.0028 0.0027	0.197	0.236	0.880	0.278	0.000	-5.883
NikkeiCac	MAGHmaxPC	0.0008 0.0007	0.0084 0.0069	0.170	0.236	0.646 0.902	0.700 0.957	0.009 0.000	-5.839
NikkeiCac	MAHmaxPC	0.0008 0.0008	0.0067 0.0058	0.352	0.236	1.000 1.000	0.564 0.859	0.008 0.000	-5.838
NikkeiCac	MAGHmaxPC sym.	0.0000 0.0003	0.0090 0.0076	0.278	0.236	0.652 0.639	0.714 1.020	0.000 0.000	-5.838
NikkeiCac	MAGHmax	0.0059 0.0018	0.0016 0.0067	0.987	0.236	0.635 0.696	0.050 0.732	-0.491 -0.015	-5.964
NikkeiCac	MAHmax	0.0044 0.0015	0.0012 0.0049	0.976	0.236	1.000 1.000	0.050 0.587	-0.322 -0.015	-6.123
NikkeiCac	MAGHmax sym.	0.0000 0.0003	0.0016 0.0065	0.986	0.236	0.636 0.689	0.050 0.709	0.000 0.000	-5.997
NikkeiCac	MGH	0.0004 0.0005	0.0032 0.0031	0.217	0.236	1.117	0.355	-0.026 -0.014	-5.885
NikkeiCac	MH	0.0004 0.0005	0.0043 0.0041	0.216	0.236	1.000	0.459	-0.028 -0.017	-5.885
NikkeiCac	MGH sym.	-0.0000 0.0003	0.0032 0.0031	0.217	0.236	1.118	0.349	0.000 0.000	-5.885

Table 5.5: Estimations from two- and four-dimensional asset-return data (NEMAX-NASDAQ, DAX-DOW-NIKKEI-CAC, DAX-NEMAX-DOW-NASDAQ).

Data	Model	Location μ	Scaling S	λ	α	β	CrossEnt
NemaxNasdaq	MAGHmin	0.0029 0.0023	0.0057 0.0044	1.562	0.337	-0.052	-4.610
NemaxNasdaq	MAHmin	0.0021 0.0018	0.0169 0.0129	1.000	0.902	-0.045	-4.609
NemaxNasdaq	MAGHmin sym.	-0.0002 0.0005	0.0011 0.0008	1.705	0.072	0.000	-4.760
NemaxNasdaq	MAGHmaxPC	-0.0000 -0.0001	0.0247 0.0151	1.143 1.111	1.255 1.542	0.007 0.000	-4.596
NemaxNasdaq	MAHmaxPC	-0.0000 -0.0000	0.0266 0.0160	1.000 1.000	1.329 1.599	0.007 0.000	-4.596
NemaxNasdaq	MAGHmaxPC sym.	-0.0006 -0.0001	0.0291 0.0172	0.797 0.837	1.424 1.669	0.000 0.000	-4.596
NemaxNasdaq	MAGHmax	-0.0013 0.0004	0.0133 0.0233	1.913 0.858	0.321 2.183	0.010 -0.020	-4.592
NemaxNasdaq	MAHmax	-0.0013 0.0005	0.0252 0.0230	1.000 1.000	1.242 2.219	0.013 -0.021	-4.592
NemaxNasdaq	MAGHmax sym.	-0.0013 -0.0003	0.0100 0.0188	2.209 1.512	0.050 1.915	0.000 0.000	-4.938
NemaxNasdaq	MGH	0.0007 0.0016	0.0040 0.0031	1.907	0.263	-0.005 -0.037	-4.616
NemaxNasdaq	MH	0.0008 0.0019	0.0214 0.0164	1.000	1.205	-0.006 -0.051	-4.614
NemaxNasdaq	MGH sym.	-0.0003 0.0003	0.0055 0.0042	1.937	0.363	0.000 0.000	-4.614
DaxDowNikkeiCac	MAGHmin	0.0009 0.0007 0.0002 0.0006	0.0051 0.0038 0.0051 0.0049	0.888	0.539	-0.020	-12.405
DaxDowNikkeiCac	MAHmin	0.0008 0.0006 0.0001 0.0005	0.0044 0.0033 0.0044 0.0042	1.000	0.481	-0.017	-12.405
DaxDowNikkeiCac	MAGHmin sym.	0.0004 0.0004 -0.0001 0.0003	0.0050 0.0037 0.0050 0.0048	0.894	0.525	0.000	-12.405
DaxDowNikkeiCac	MAGHmaxPC	0.0016 0.0015 0.0005 0.0010	0.0099 0.0047 0.0022 0.0045	0.92 0.90 1.05 0.83	0.31 0.17 0.52 1.33	0.032 0.000 0.000 0.000	-12.403
DaxDowNikkeiCac	MAHmaxPC	-1.5272 -0.7592 -0.8804 -1.4177	0.0093 0.0062 0.0016 0.0041	1.00 1.00 1.00 1.00	0.05 0.05 0.79 1.29	0.039 0.000 0.000 0.000	-12.827
DaxDowNikkeiCac	MAGHmaxPC sym.	0.0009 0.0007 0.0002 0.0008	0.0104 0.0045 0.0022 0.0048	0.63 0.88 1.01 0.63	0.05 0.16 0.46 1.40	0.000 0.000 0.000 0.000	-12.451
DaxDowNikkeiCac	MAGHmax	0.0010 0.0013 0.0049 0.0003	0.0085 0.0040 0.0007 0.0077	0.87 1.09 0.63 0.72	0.24 0.61 0.05 1.37	-0.039 -0.051 -0.464 0.054	-12.505
DaxDowNikkeiCac	MAHmax	-0.9333 -0.2867 -0.2327 -0.6268	0.0073 0.0033 0.0007 0.0070	1.00 1.00 1.00 1.00	0.05 0.47 0.05 1.30	0.600 0.007 -0.100 0.046	-12.869
DaxDowNikkeiCac	MAGHmax sym.	0.0001 0.0004 0.0000 0.0000	0.0085 0.0030 0.0008 0.0082	0.63 0.98 0.63 0.637	0.05 0.39 0.05 1.42	0.000 0.000 0.000 0.000	-12.575
DaxDowNikkeiCac	MGH	0.0013 0.0009 0.0004 0.0007	0.0046 0.0036 0.0049 0.0045	2.000	0.726	-0.032 -0.023 -0.021 -0.019	-12.441
DaxDowNikkeiCac	MH	0.0012 0.0009 0.0004 0.0007	0.0076 0.0059 0.0081 0.0075	1.000	0.939	-0.035 -0.026 -0.023 -0.023	-12.453
DaxDowNikkeiCac	MGH sym.	0.0005 0.0005 -0.0001 0.0003	0.0043 0.0033 0.0045 0.0042	2.000	0.664	0.000 0.000 0.000 0.000	-12.441
DaxNemDowNasd	MAGHmin	0.0006 0.0006 0.0006 0.0009	0.0104 0.0177 0.0076 0.0136	1.136	1.000	-0.015	-10.903
DaxNemDowNasd	MAHmin	0.0006 0.0005 0.0006 0.0009	0.0115 0.0197 0.0084 0.0151	1.000	1.087	-0.015	-10.903
DaxNemDowNasd	MAGHmin sym.	-0.0001 -0.0004 0.0002 0.0004	0.0106 0.0180 0.0077 0.0138	1.127	1.019	0.000	-10.903
DaxNemDowNasd	MAGHmaxPC	0.0011 0.0008 -0.0001 -0.0000	0.0357 0.0203 0.0081 0.0037	1.63 0.84 1.85 0.90	1.08 1.89 0.26 2.18	-0.030 0.000 0.000 0.000	-10.868
DaxNemDowNasd	MAHmaxPC	0.0012 0.0005 -0.0002 -0.0002	0.0358 0.0201 0.0105 0.0062	1.00 1.00 1.00 1.00	1.47 1.84 1.14 2.13	-0.035 0.000 0.000 0.000	-10.868
DaxNemDowNasd	MAGHmaxPC sym.	-0.0000 -0.0005 0.0001 -0.0002	0.0377 0.0207 0.0092 0.0038	1.40 0.65 2.18 0.81	1.29 1.95 0.05 2.14	0.000 0.000 0.000 0.000	-11.217
DaxNemDowNasd	MAGHmax	0.0016 0.0006 0.0010 0.0027	0.0145 0.0023 0.0127 0.0072	1.16 1.98 0.64 1.02	1.19 0.05 1.79 0.70	-0.060 0.010 -0.018 -0.066	-11.279
DaxNemDowNasd	MAHmax	0.0017 0.0006 0.0010 0.0027	0.0162 0.0060 0.0119 0.0073	1.00 1.00 1.00 1.00	1.27 0.37 1.77 0.72	-0.063 0.029 -0.013 -0.067	-10.939
DaxNemDowNasd	MAGHmax sym.	0.0000 -0.0012 0.0002 0.0011	0.0172 0.0006 0.0118 0.0003	0.65 1.89 0.64 2.73	1.42 0.05 1.84 0.06	0.000 0.000 0.000 0.000	-11.583
DaxNemDowNasd	MGH	0.0014 0.0021 0.0011 0.0023	0.0038 0.0064 0.0028 0.0049	1.990	0.434	-0.014 -0.026 -0.003 -0.053	-10.967
DaxNemDowNasd	MH	0.0015 0.0022 0.0010 0.0024	0.0132 0.0220 0.0095 0.0169	1.000	1.258	-0.019 -0.030 -0.001 -0.064	-10.964
DaxNemDowNasd	MGH sym.	0.0001 -0.0001 0.0003 0.0005	0.0027 0.0044 0.0019 0.0034	1.987	0.296	0.000 0.000 0.000 0.000	-10.968

5.2 Semi-parametric model

Recall from Section 1 that standard portfolio selection is based on the Markowitz mean-variance theory and the Sharpe-Lintner-Mossin capital asset pricing model (CAPM). Here the risk-return optimum is determined by the covariance matrix Σ and the mean vector μ of the corresponding asset-return vector. The Markowitz theory and the CAPM are well established within the world of multivariate normal distributed asset-return vectors. However, as we have already mentioned before, the multivariate normal distribution has been questioned in the context of financial asset-returns on two principal reasons: empirical evidence indicates that most financial data show both pronounced asymmetry and much heavier tail behavior than is consistent with normality. Given the importance of appropriate tail modelling in the context of modern risk management based on risk measures such as Value-at-Risk this drawback became even more relevant. Furthermore the role of the covariance matrix as the only measure of dependence was already shown to be insufficient.

Thus, there has been much interest in developing financial models which preserve at least some of the advantages of the classical normal distribution - familiarity, tractability, interpretability of mean vector μ and covariance matrix Σ - without the drawbacks of being unable to handle asymmetry or heavy tails - *skewness* and *kurtosis*. However, the class of normal distributions remains, of course, extremely important for asset-return modelling, as a benchmark against which any alternative distribution must first be judged.

5.2.1 Normal variance-mean mixtures and elliptical distributions

Normal variance-mean mixtures

Various parametric models have been proposed in mathematical finance for asset-return modelling - in particular the stable and hyperbolic models (cf. Section 5.1). However, reality is always more complicated than can be captured by just a few parameters, and so a nonparametric approach may also be adopted. The intention of this section is to show that the right way is to combine the advantages of the parametric and nonparametric approaches, using a *semi-parametric* model. In particular, we use a *parametric* component, incorporating the mean vector μ and covariance matrix Σ, and a *nonparametric* component, modelling the *shape* of the distribution - specifically, questions of tail-decay (kurtosis). Here, shape - which we can think of as a density on $[0, \infty)$ - incorporates what remains when we work up to location and scatter - that is, modulo affine transformations - while (μ, Σ) represents the affine part.

In the following section we mainly concentrate on elliptically contoured distributions introduced in Section 3.2 (cf. also Bickel, Klaasen, Ritov, and Wellner (1998)). Most of the results were elaborated in Bingham, Kiesel, and Schmidt (2002) which serves as the basis of this section. To begin with we summarize the theoretical background, largely following Bingham and Kiesel (2002).

One possible framework in which we can escape the limitations of the normal distributions while still retaining some of the convenience of the latter class of distributions is that of *normal variance-mean mixtures* (NVMM)), for a good review we refer to Barndorff-Nielsen, Kent, and Sørensen (1982a). Let Z be a random variable with some distribution function F living on $[0,\infty)$. If the n-dimensional conditional random vector $X|(Z=z)$ is multivariate normal distributed, i.e.,

$$X|(Z=z) \sim N_n(\mu + z\beta, z\Sigma) \tag{5.34}$$

- where $\mu, \beta \in I\!\!R^n$ and $\Sigma \in I\!\!R^{n\times n}$ with determinant one - then X is called a *normal variance-mean mixture* with *mixing distribution function* F. In terms of the Laplace-Stieltjes transform Φ of F, that is

$$\Phi(s) := \int_0^\infty e^{-sz} dF(z), \quad s > 0,$$

the characteristic function ψ_X of X can be written as

$$\psi_X(t) = \exp\left\{it'\mu\right\} \Phi\left(\frac{1}{2}t'\Sigma t - it'\beta\right). \tag{5.35}$$

Normal variance mixtures

If we specialize to $\beta = 0$ in the above, we obtain the class of *normal variance mixtures* (NVM):

$$X|(Z=z) \sim N_n(\mu, z\Sigma); \quad \psi_X(t) = \exp\left\{it'\mu\right\} \Phi\left(\frac{1}{2}t'\Sigma t\right). \tag{5.36}$$

In this case the distribution of X is isotropic, and we may - and do - drop the restriction to unit determinant, imposed above for reasons of identifiability.

The class NVM of normal variance mixtures has important advantages over NVMM. One that concerns us is that structural information on the mixing distribution function F transfers to the mixture distribution. For example, if F is *infinitely divisible* ($F \in ID$), then the distribution of X possesses this property too, as follows immediately from (5.36).

Subclasses of ID are also important here. The distribution of X belongs to the class of *self-decomposable* distributions (SD) if for each $\rho \in (0,1)$ there exists a characteristic function ψ_ρ for which

$$\psi_X(t) = \psi_X(\rho t) \cdot \psi_\rho(t). \tag{SD}$$

Self-decomposable distributions are infinitely divisible, i.e.,

$$SD \subset ID,$$

and self-decomposability of F again transfers to self-decomposability of (the distribution of) X. We write $SDNVM$ for the resulting class of self-decomposable normal variance mixtures.

Examples.
1. *Multivariate normal distribution.* The distribution $N_n(\mu, \Sigma)$ is in SDNVM, with

$$\Phi(u) = \exp(-u/2).$$

2. *Generalized hyperbolic distribution.* A particular choice of the mixing distribution, in particular the *generalized inverse Gaussian* distributions (GIG) which is a three-parameter family, gives rise to the *generalized hyperbolic* distributions (GH). Here Φ involves the quotient of Bessel functions K_λ of the third kind. For details, background and references, see Bingham and Kiesel (2001b). The generalized hyperbolic distributions have been also discussed in Section 5.1.

The class GH, or more generally NVM, has been used in higher dimensions in the context of portfolio theory by Bingham and Kiesel (2001a) or Eberlein (2001) who utilize the theory of elliptically contoured distributions (see below). The latter class is more tractable but less general than NVMM. By contrast, Korsholm (2000) used NVMM for financial modelling in one dimension, where it was possible to handle asymmetry.

Elliptically contoured distributions

Recall that an n-dimensional distribution is spherical if it is invariant under the orthogonal rotation group $O(n)$. Further, a distribution belongs to the family of elliptically contoured distributions (E) if it is the image of a spherical distribution under an affine transformation. For convenience, we confine attention to absolutely continuous elliptically distributed random vectors $X \in E_n(\mu, \Sigma, g)$ which are in the L_2 space and have non-degenerated margins. Then the mean vector μ and covariance matrix Σ exist; Σ is invertible; the density $f(x)$ exists, and is a function of the quadratic form

$$Q(x) := (x - \mu')\Sigma^{-1}(x - \mu); \text{ i.e.,} \quad f(x) = |\Sigma|^{-\frac{1}{2}} g(Q) \tag{5.37}$$

for some function $g : I\!\!R_+ \to I\!\!R_+$ which we call the density generator. Recall from Section 3.2 that the matrix Σ is uniquely determined up to a positive constant and in the L_2 space w.l.o.g. it corresponds to the covariance matrix. Further, (μ, Σ) is the parametric part of our distribution model (mean and covariance, as in the Markovitz theory), while g is the nonparametric part, and contains information on the 'shape' of the distribution - tail decay, etc. According to Section 3.2, the definition of elliptically contoured distributions does not require the distribution to belong to the L_2 or even L_1 space. In that case Σ denotes some scaling matrix and μ is some location vector.

Recall that the characteristic function has the form

$$\psi_X(t) = \exp\left\{it'\mu\right\} \cdot \Phi\left(t'\Sigma t\right)$$

with Φ called the characteristic generator. For more background, see Fang, Kotz, and Ng (1990b), or Bingham and Kiesel (2002), Chapter 2 (see also Section 3.2 of the present section).

Self-decomposable elliptically contoured distributions

There are a number of theoretical advantages in combining the properties of self-decomposability with being elliptically contoured: call the resulting class SDE.

(i) Self-decomposable distributions are absolutely continuous and unimodal (Sato (1999), Theorems 28.4 and 53.1). So for SDE, the density generator *g exists and is decreasing*. This provides a far-reaching and easily visualized generalization of the Gaussian case, where

$$g(u) = \exp(-u/2)/(2\pi)^{\frac{1}{2}n}.$$

It also allows us to incorporate the structural information in our estimation procedure (see Section 5.2.2).

(ii) Self-decomposability is well adapted to modelling the *dynamic* (time-series) aspects of portfolios, as well as the *static* (distributional) aspects above. For, the self-decomposability property corresponds well to autoregressive time-series models

$$X_t = \rho X_{t-1} + \varepsilon_t$$

with ε_t the innovation or error term. This link is developed , e.g., in Barndorff-Nielsen, Jensen, and Sørensen (1998).

(iii) In addition, for the normal variance-mixture case, self-decomposability transfers from the mixing distribution to the mixture distribution, as noted above:

$$NVM \subset E, \quad SDNVM \subset SDE.$$

Examples. 1. The multivariate symmetric ($\beta = 0$) generalized hyperbolic distributions (MSGH) are self-decomposable (Halgreen (1979)) and elliptically contoured:

$$MSGH \subset SDE.$$

See e.g. Bingham and Kiesel (2001b) for background and references (cf. also Section 5.1 of the present thesis).
2. The *variance gamma* distributions are also in SDE. For background, see Madan and Seneta (1990), Carr, Chang, and Madan (1998).
3. The multivariate t-distributions are in SDE; see Fang, Kotz, and Ng (1990b), Section 3.3.6.

5.2.2 Estimation

Fitting data to an elliptically contoured distribution requires estimation of the parametric location and scatter parameters (μ, Σ) and the nonparametric shape function g. We present robust and appropriate estimators fulfilling the prerequisite of a decreasing density generator g. Consider a random sample $X^{(1)}, \ldots, X^{(m)}, \ m \in \mathbb{N}$.

Estimating the mean and the covariance

Natural estimators for the mean vector μ and the covariance matrix Σ of a multivariate elliptically contoured distribution are the sample mean and the Pearson product-moment covariance. Although the latter estimator almost surely yields a positive-definite covariance matrix, which avoids unpleasant matrix transformations, both estimators are vulnerable to heavy-tailed and contaminated data: a single grossly aberrant reading can destroy the accuracy of both. Several robust multivariate estimation techniques for μ and Σ have been proposed in the literature. Among them we mention:

(i) Multivariate trimming (Hahn, Mason, and Weiner 1991),

(ii) Minimum-volume ellipsoid estimator for Σ (Rousseeuw and van Zomeren 1990),

(iii) Median estimator for μ and minimum-angle estimator for Σ introduced in Section 4.9.1.

In Section 4.9.1 we have already shown that the third set of estimators has much to recommend it. First, they provide robust estimators utilizing exactly the structure of elliptically contoured distributions. Second, they are applicable even if the covariance matrix or the mean vector do not exist. Recall that the estimate obtained in Section 4.9.1, in particular $\hat{\Sigma}_* = \hat{A}'_*\hat{A}_*$, is not necessarily a covariance matrix but might be some scaling matrix. In order to obtain an estimate $\hat{\Sigma} = \hat{A}'\hat{A}$ of the covariance matrix Σ, if it exists, we must rescale $\hat{\Sigma}_*$ by $\hat{\mu}_{R_*^2}/2$ with $\hat{\mu}_{R_*^2}$ being some estimator of $E(R_*^2) = E(||(\hat{A}'_*)^{-1}(X - \hat{\mu})||_2)$.

Note that approximate realizations of the squared generating variate R_n^2 are obtained by $(x_i - \hat{\mu})'\hat{\Sigma}^{-1}(x_i - \hat{\mu})$, $i = 1, \ldots, m$. These data are exploited for the density generator estimation in the next section.

Estimating the density generator g

Two estimation methodologies are distinguished for the density generator of elliptically contoured distributions. Either the density generator is given in parametric form, as in the context of hyperbolic or t-distributions, or the generator is assumed to be nonparametric.

For both estimation approaches we utilize the transformed data

$$q_i = (x_i - \hat{\mu})'\hat{\Sigma}^{-1}(x_i - \hat{\mu}), \quad i = 1, \ldots, m$$

where x_i is a realization of $X^{(i)}$, $i = 1, \ldots, m$. The data points q_i are approximate realizations of the underlying distribution given by the squared random variate R_n^2. In particular the latter data come from a density $f_{R_n^2}$ which relates to the unknown density generator g as follows (cf. Fang, Kotz, and Ng (1990b), Remark to Theorem 2.9, or Lemma 3.1.12 in Section 3.1.1)

$$f_{R_n^2}(u) = \frac{\pi^{n/2}}{\Gamma(n/2)} u^{n/2-1} g(u). \tag{5.38}$$

Nonparametric estimation. For nonparametric estimation of the density $f_{R_n^2}$ given in (5.38), we use standard results on optimal bandwidth selection, see e.g. Härdle (1990b), Chapter 4, or Wand and Jones (1995), Chapter 3. In particular, we use the variants of direct plug-in rules (Wand and Jones (1995), Section 3.6.).

Let h be some arbitrary density function which is unknown. According to the above references, the asymptotic integrated mean-squared error (A-MISE) optimal bandwidth for kernel density estimation is given by

$$b_0 = \left(\frac{\|K\|_2^2}{\|h''\|_2^2 \, (\mu_2(K))^2 m} \right)^{\frac{1}{5}}$$

(here K is a kernel, $\|.\|_2$ is the Euclidean-norm, and $\mu_2(K) = \int x^2 K(x)dx$). It turns out to be problematic to deal with the second derivative of the unknown density h. A first idea is to replace h'' by the corresponding value of a reference distribution. For our applications it suffices to use the Gaussian kernel and the normal distribution as reference distribution (cf. Härdle (1990b)). However, in many contexts of financial data we are often confronted with heavier tailed distributions. Therefore, a careful choice of the kernel and the reference distribution might be necessary to make the bandwidth insensitive to outliers.

A sophistication of the above idea is given by multi-stage plug-in estimators: Instead of utilizing a reference distribution, a kernel estimate of the density functional $\|h''\|_2^2$ is used, which will again be sensitive to bandwidth choice. Motivated by our simulations and detailed comparison studies of data-driven bandwidth selection principles (see e.g. Park and Marron (1990) and Wand and Jones (1995), Section 3.8., for data from heavy-tailed distributions) we employed a two-stage plug-in estimator suggested by Sheather and Jones (1991) and outlined in Wand and Jones (1995), Section 3.6.

Similar problems regarding the choice of bandwidth arise while estimating confidence bands. Asymptotic confidence bands for the latter nonparametric estimation of the density h are given by

$$\left[\hat{h}_b(x) - m^{-\frac{1}{5}} \left(\frac{c^2}{2} h''(x)\mu_2(K) + d_\alpha \right), \hat{h}_b(x) - m^{-\frac{1}{5}} \left(\frac{c^2}{2} h''(x)\mu_2(K) - d_\alpha \right) \right],$$

where $d_\alpha = u_{1-\alpha/2}\sqrt{c^{-1}h(x)\,\|K\|_2^2}$, $u_{1-\alpha/2}$ is the $1-\alpha/2$-quantile of the standard normal distribution, and c is obtained from the relationship $b_0 = cm^{-\frac{1}{5}}$.

Note that density generators have non-negative support which requires a density estimation near the boundary points. This causes a bad performance of the latter estimation methods. Several possible modifications to improve the performance near the boundaries are discussed in Härdle (1990a), Section 4.4, Wand and Jones (1995), Section 2.11, Müller (1991), and Jones (1993). In the context of elliptically contoured distributions the problem is discussed in Stute and Werner (1991) and Hodgson, Linton, and Vorkink (2000). Similar to the latter references, our simulation results confirm that a modification suggested by Schuster (1985) works well. In our setting, Schuster's modification incorporates the additional information of positive support of the underlying density by adding a mirror image term to the standard kernel estimator. In particular the following estimator for the density function h is utilized

$$\hat{h}_m(x) = \frac{1}{m\hat{b}_0} \sum_{i=1}^{m} \left[K\left(\frac{x - x_i}{\hat{b}_0}\right) + K\left(\frac{x + x_i}{\hat{b}_0}\right) \right].$$

According to Section 5.2.1, one of the main structural features we require, is the monotonicity of the density generator g. For that we transform the estimated density generator via a monotone regression method suggested in Härdle (1990a), Section 8.1. However, this regression method assumes an unimodal density function which has to be verified (compare also Anderson, Fang, and Hsu (1986) or Fang and Zhang (1990), Section 4.1). The reason for unimodality is because we mirror the observations on a possible mode point and then use the monotone regression algorithm of the partially mirrored data.

Parametric estimation. If the density generator g is assumed to be of a parametric form, we may estimate the parameters of the corresponding density $f_{R_n^2}$ (see equation (5.38)) via maximum-likelihood techniques, utilizing the approximate realizations $q_i = ||\hat{\Sigma}^{-1/2}(x_i - \hat{\mu})||_2,\ \ i = 1, \ldots, m$. This approach might not be satisfactory in the case of VaR considerations, where we are particularly interested in the tail behavior of the density generator. The maximum-likelihood estimate yields an overall fit including all available data and usually underestimates the tail, whereas VaR primarily examines the tail behavior and tail dependence of a multivariate portfolio distribution. Assuming a regularly-varying density generator we propose estimating the associated tail parameters via estimates of the tail index or the tail-dependence coefficient. To give an example we consider a multivariate t-distribution with $\nu > 2$ degrees of freedom and density generator

$$g(u) = \frac{1}{((\nu - 2)\pi)^{\frac{n}{2}}} \frac{\Gamma((\nu + n)/2)}{\Gamma(\nu/2)} \left(1 + \frac{u}{\nu - 2}\right)^{-\frac{1}{2}(n+\nu)}, \quad u > 0.$$

Note that the density generator possesses a slightly different representation for $\nu \leq 2$. If $\hat{\alpha},\ \hat{\lambda}$ and $\hat{\rho}$ denote some estimators for the tail-index, the tail-dependence coefficient matrix $\lambda = (\lambda_{ij})$, and the correlation coefficient matrix $\rho = (\rho_{ij})$, respectively, then an estimate for ν can be obtained via the relation $2 + \alpha = n + \nu$, so $\alpha = \nu$ for $n = 2$, and formula (3.18) (for the numerics, note that the right of (3.18) is monotone in α).

5.2.3 Simulation

In this section we address the generation of multivariate pseudo random vectors from a multivariate elliptically contoured distribution. Efficient and fast multivariate random-number generators are indispensable for modern portfolio theory. Monte-Carlo simulations for Value-at-Risk and pricing calculations require many samples, which have to be sampled in a reasonable time-frame. For the class of elliptical distributions efficient and fast algorithms can be presented. Precisely, for elliptical distributions the multidimensional simulation problem boils down to a one-dimensional simulation task. In the following context of random number generation we assume that Σ is a positive-definite matrix, such that the Cholesky decomposition $\Sigma = A'A$ results in a unique full-rank lower-triangular matrix $A' \in I\!R^{n \times n}$. The connection between the density generator g and the density function $f_{R_n^2}$ of the random variable R_n^2 is stated in formula (5.38). A generic algorithm for generating multivariate elliptically-distributed random vectors is given below and is based on the stochastic representation (3.16). For elliptical distributions the algorithm becomes faster the easier the random variate R_n is simulated.

Pseudo algorithm for generating multivariate elliptical random numbers from $E_n(\mu, \Sigma, g)$ distributions:

1. Set $\Sigma = A'A$ via Cholesky decomposition.
2. Sample a random number from R_n.
3. Sample n independent random numbers $S_1, \dots, S_n$ from an univariate standard-normal distribution $N_1(0, 1)$.
4. Set $S = (S_1, \dots, S_n)$.
5. Set $U^{(n)} = ||S||^{-1} \cdot S$.
6. Return $X = \mu + R_n \cdot A'U^{(n)}$.

Simulating from a parametric elliptically contoured distribution

The above generic simulation algorithm serves as an universal tool for the entire class of elliptical distributions. However, below we present special generation algorithms based on different mixture representations for two well-known families of elliptical distributions. Here Σ corresponds to the covariance matrix and the algorithms are formulated accordingly.

Pseudo algorithm for generating multivariate random numbers from a multivariate t-distribution $Mt_n(\mu, \Sigma, \nu),\ \nu > 2$:

1. Set $\Sigma = A'A$ via Cholesky decomposition.
2. Sample a random number T from χ_ν.
3. Sample n independent random numbers $S_1, \dots, S_n$ from an univariate standard-normal distribution $N_1(0, 1)$.
4. Set $S = (S_1, \dots, S_n)$.
5. Return $X = \mu + \frac{\sqrt{\nu-2}}{T} \cdot A'S$.

Pseudo algorithm for generating multivariate random numbers from a multivariate generalized hyperbolic distribution $MGH_n(\mu, \Sigma, (\lambda, \psi))$:

1. Set $\Sigma = A'A$ via Cholesky decomposition.

2. Sample an inverse Gaussian random number Y with parameters λ, ψ and χ.

3. Sample n independent random numbers $S_1, \ldots, S_n$ from an univariate standard-normal distribution $N_1(0, 1)$.

4. Set $S = (S_1, \ldots, S_n)$.

5. Return $X = \mu + \sqrt{Y} A'S/c$ (The constant is chosen such that Σ represents the covariance matric (cf. Theorem 5.1.5))

For the density generator of the generalized inverse Gaussian (GIG) distribution and a detailed account of an efficient simulation algorithm we refer to Section 5.1.1.

Simulating from a semi-parametric elliptically contoured distribution

Simulation from a semi-parametric $E_n(\mu, \Sigma, \hat{g})$ distribution utilizes formula (5.38). This formula yields a density generator estimate $\hat{g}$ which is obtained from the estimated density function $\hat{f}_{R_n^2}$. The latter density is estimated via the nonparametric estimation methods described in Section 5.2.2. The tail behavior of $\hat{g}$ is modelled via a power law in order to include tail dependence (see Theorem 3.2.2). This leads to a straightforward random-number generator based on the well-known inversion transformation method (see, for example, Niederreiter (1993)).

5.2.4 Application to financial data

The financial time series under consideration consist of 6 series of major stock indices covering the (sub-) periods between January 1987 to December 2002 (data obtained from Bloomberg Financial Services). We investigate the density generators of various portfolios of indices. In particular we use

- Mature markets (from 1987 to 2002): DOW JONES, DAX, CAC40, NIKKEI;

- New economy markets (from 1998 to 2002): NASDAQ, NEMAX.

In addition, we use 6 major US stocks, namely Ford, Boeing, General Motors, Dell Computers, Cisco Systems, Microsoft during the period 1990 to 2002 to investigate portfolios of stocks.

Fitting the data

We start with several visual tests for elliptical symmetry of the underlying distribution. In particular, we provide plots of the four time-series (Figure 5.6), contour and relief plots of the estimate of the density (Figure 5.7) (obtained by using the function kde2d from Venables and Ripley (1999) in S-Plus), and the estimate of the density generator g (Figure 5.8, 5.9, 5.10 and 5.12).

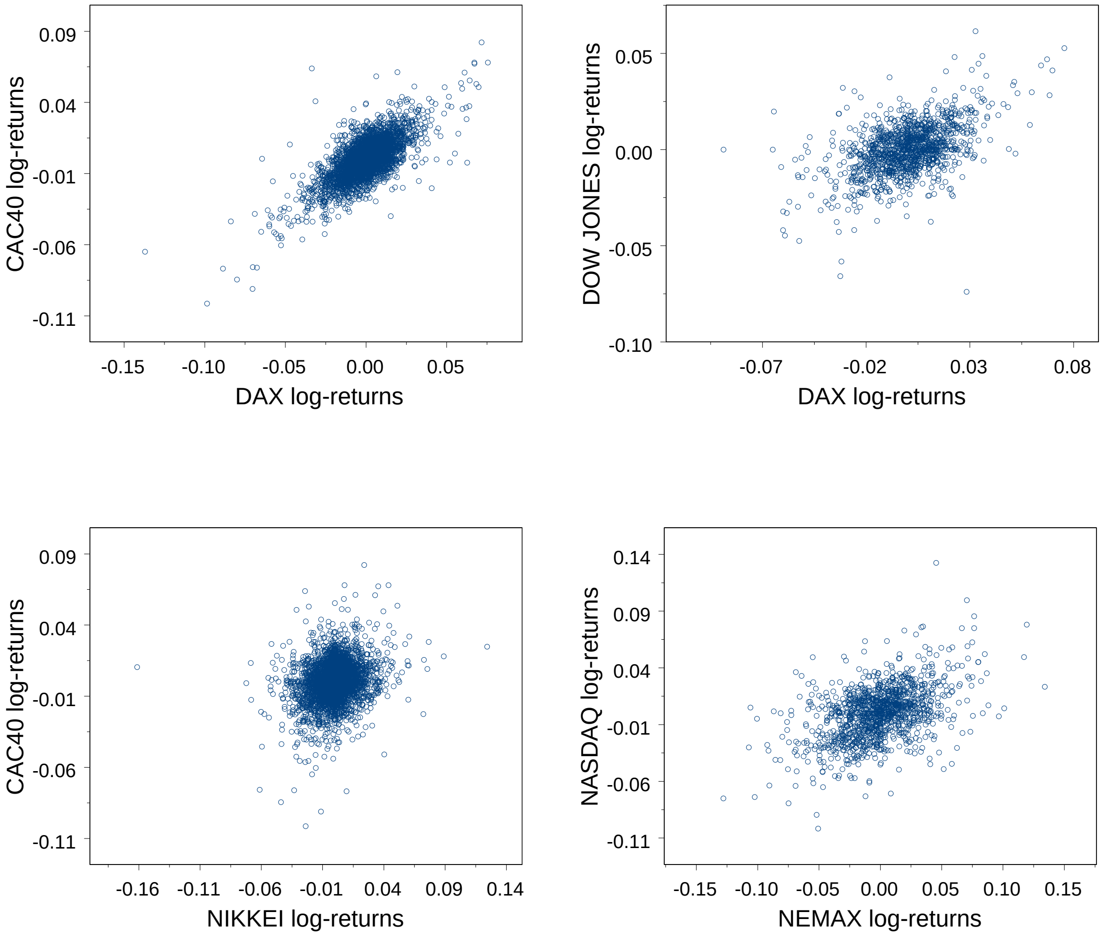


Figure 5.6: Scatter plots of bivariate log-returns of four asset series.

In $n = 2$ dimensions, where the density is easiest to visualize, the output from applying our procedure to a data set – real or simulated – is a relief and/or contour-plot of the density, and a plot of the estimated density generator g. In the general case, we cannot visualize the density f so easily (see Scott (1992) for what can be done here), and our principal output is the plot of g. There are two broad uses to which our g-plot can be put. The first is to test goodness of fit to a model. The second is to estimate the tail-decay of g, which reflects the tail-decay of the data itself, one of the prime features under investigation here.

The density generator g focuses on properties of the portfolio as a whole, while (μ, Σ) focuses on properties of the individual assets within the portfolio. Thus, if

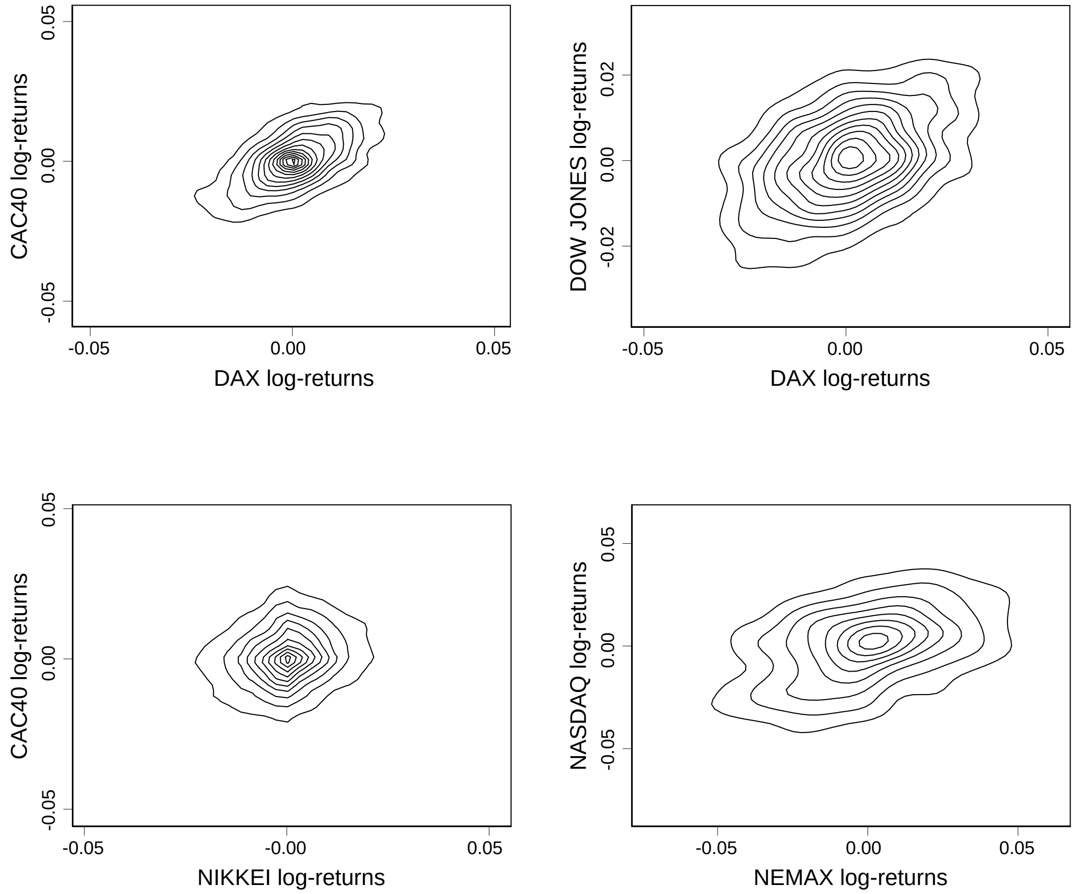


Figure 5.7: Contour-plots of bivariate log-returns of four asset series.

we wish to rank our preferences between the individual assets in a fixed basket of assets, our preferences are the same in any fixed elliptically contoured world as in the normal/Gaussian world (since (μ, Σ) characterize any coherent risk measures in such settings, see Embrechts, McNeil, and Straumann (2001), Section 3.4.). By contrast, if we compare the density generator g of a typical 'old economy' (of Germany's DAX, the USA's DOW JONES, etc.) with that of a typical 'new economy' portfolio (of Germany's NEMAX (Neuer Markt), the USA's NASDAQ etc.), the extra riskiness of the latter should be reflected in the fatter tail of g (as well as in an increased volatility).

Regarding time series of financial asset-returns, the distributional aspects, and tail behavior in particular, depend on the frequency with which returns are calculated (see e.g. Pagan (1996)). Since returns over long periods are sums of returns over shorter periods, the central limit theorem will give a tendency towards normality – 'aggregational Gaussianity'. This is indeed observed, for returns over longer periods than about 16 trading days – monthly returns, say – and accords with benchmark Black-Scholes-Merton theory. At the other extreme, high-frequency data – returns over the order of minutes, say – display power-law decay, for reasons involving self-similarity and scaling arguments as in physics; see e.g. Dacorogna, Müller, and Pictet (1998) and Schmitt, Schertzer, and Lovejoy (1999) for background. Intermediate between these are, say, daily returns, for which the log-linear tails of the hyperbolic/NIG model are often used, etc.

For our two-dimensional examples we used the DAX-DOW JONES, DAX-CAC40, NIKKEI-CAC40, and NEMAX-NASDAQ series; however, the results are representative for all combinations we tried.

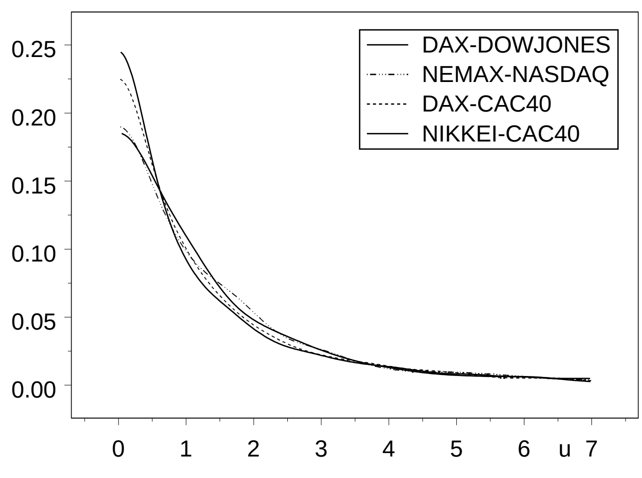


Figure 5.8: Density generators $g_i(u)$ of bivariate log-returns of four asset time series.

A more formal investigation is the use of QQ-plots as discussed in the next section.

Figure 5.11 displays such plots for the above index combinations. All plots indicate a satisfactory fit of the underlying model.

Finally we proceeded to estimate density generators using our favorite two-step selection principle. Plots for two- and higher-dimensional portfolios are displayed in Figures 5.9, 5.10 and 5.12.

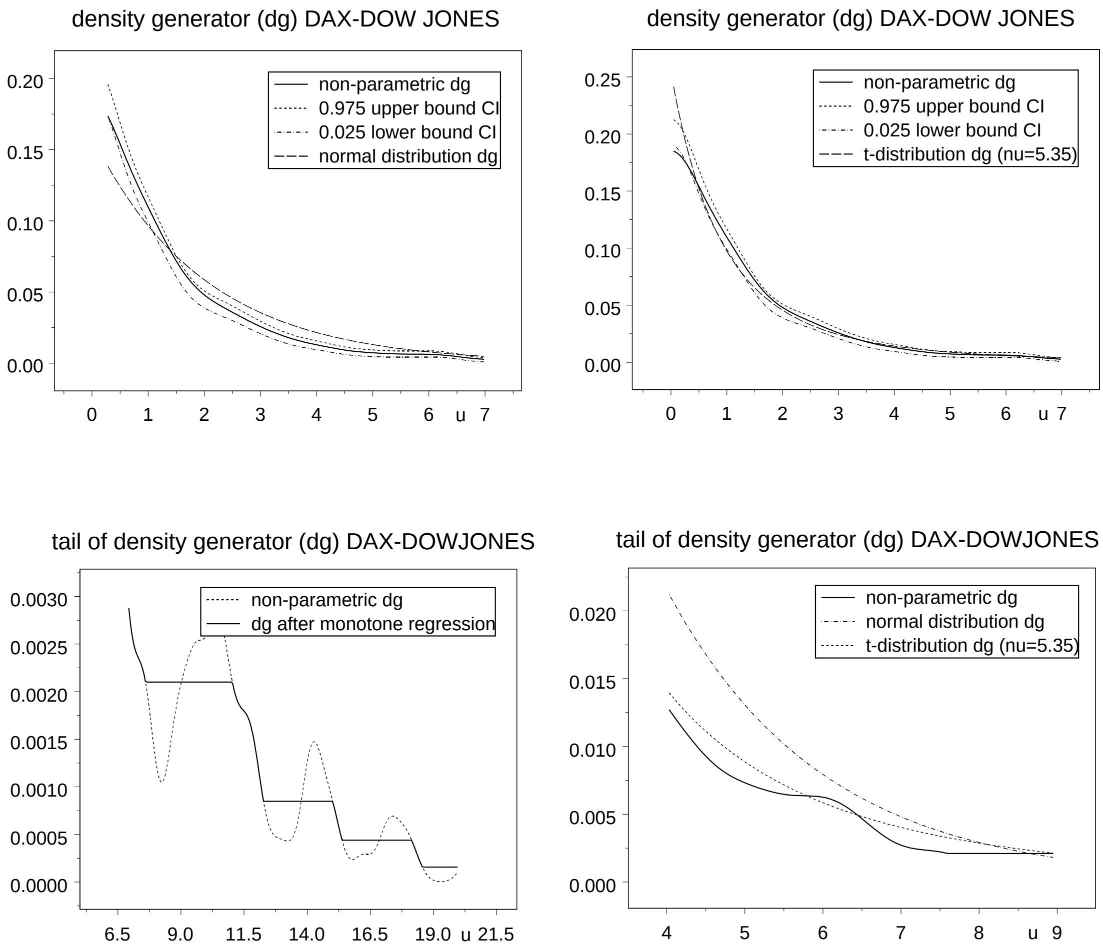


Figure 5.9: Density generators $g_i(u)$ of bivariate log-returns of DAX-DOW JONES from 1998 until 2002.

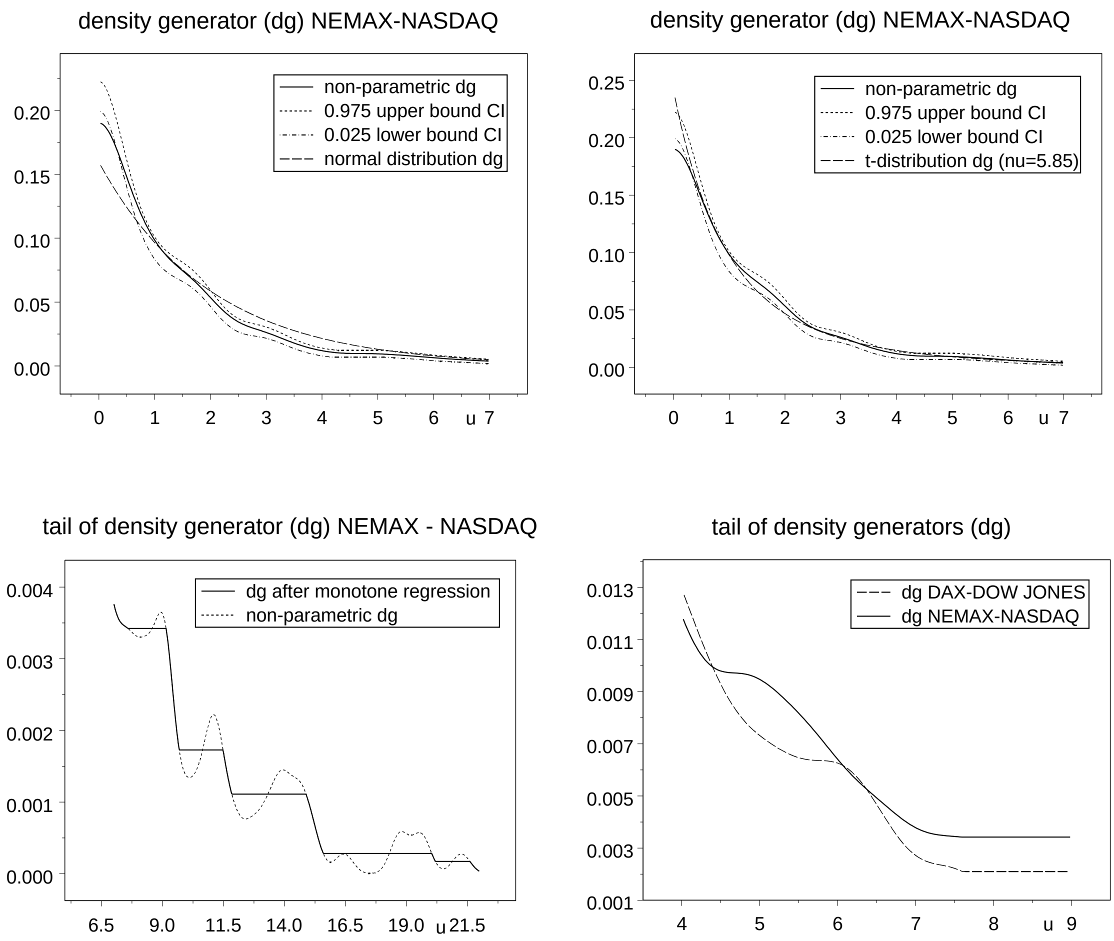


Figure 5.10: Density generators $g_i(u)$ of bivariate log-returns of NEMAX-NASDAQ from 1998 until 2002.

Tests for elliptical symmetry

To test for spherically symmetric and elliptically symmetric distributions we use a simple graphical method suggested by Li, Fang, and Zhu (1997). Alternatively, one could also use one of the tests suggested in e.g. Baringhaus (1991), Beran (1979) or Manzotti, Pérez, and Quiroz (2002).

This method is based on use of a statistic $t(x)$ which is invariant under orthogonal transformations. From Fang, Kotz, and Ng (1990b), Theorem 2.22, we know that the distribution of such a statistic remains invariant for $X \sim S_n^+(\Phi)$, i.e., n-dimensional spherical distributions with no probability mass at the origin.

For our purpose we use the standard t-statistic: For a one-dimensional random sample $X_1, X_2, \ldots, X_n$ define $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$ and $s^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$; then

$$t(X_1, X_2, \ldots, X_n) = \frac{\sqrt{n}\bar{X}}{s} \sim t_{n-1}$$

if $(X_1, X_2, \ldots, X_n) \sim N(0, I_n)$. By the above invariance principle one can replace the multivariate normal with any appropriate spherical distribution, and indeed, after normalisation, elliptical distribution. Therefore a QQ-plot of a sample against the appropriate t-distribution should indicate possible deviations from the class of spherical (ellipical) distributions (see Figure 5.11).

Further possible graphical estimators are outlined in Li, Fang, and Zhu (1997), with underlying theory given in Fang, Kotz, and Ng (1990b), Section 2.7.

Multidimensional portfolios

In Figure 5.12 several multidimensional portfolios are considered in order to stress the applicability of the semi-parametric model for higher dimensional asset-return modelling. First we construct two four-dimensional asset-return vectors consisting of the indices DAX - DOWJONES - NIKKEI - CAC40 and DAX - NEMAX - DOWJONES - NASDAQ. The first being a pure 'old' economy portfolio and the second a mixture of 'new' and 'old' economy indices. We provide the QQ-plots, introduced in Section 5.2.4, to justify applying the semi-parametric model and the corresponding density generator. Again we observe that the density generator of the t-distribution fits the nonparametric density generator much better than the normal distribution. However, the density generator of the t-distribution decreases as a power-law in the entire range of the data which is actually not observed in the data. Finally we construct a six-dimensional portfolio consisting of the stocks: FORD - BOEING - GM - DELL - CISCO - MICROSOFT. Here we point out that the semi-parametric model we propose has much to recommend it for higher-dimensional portfolio modelling. For the data we investigate that the curse of dimensionality is avoided in a satisfactory way through the nonparametric density generator.

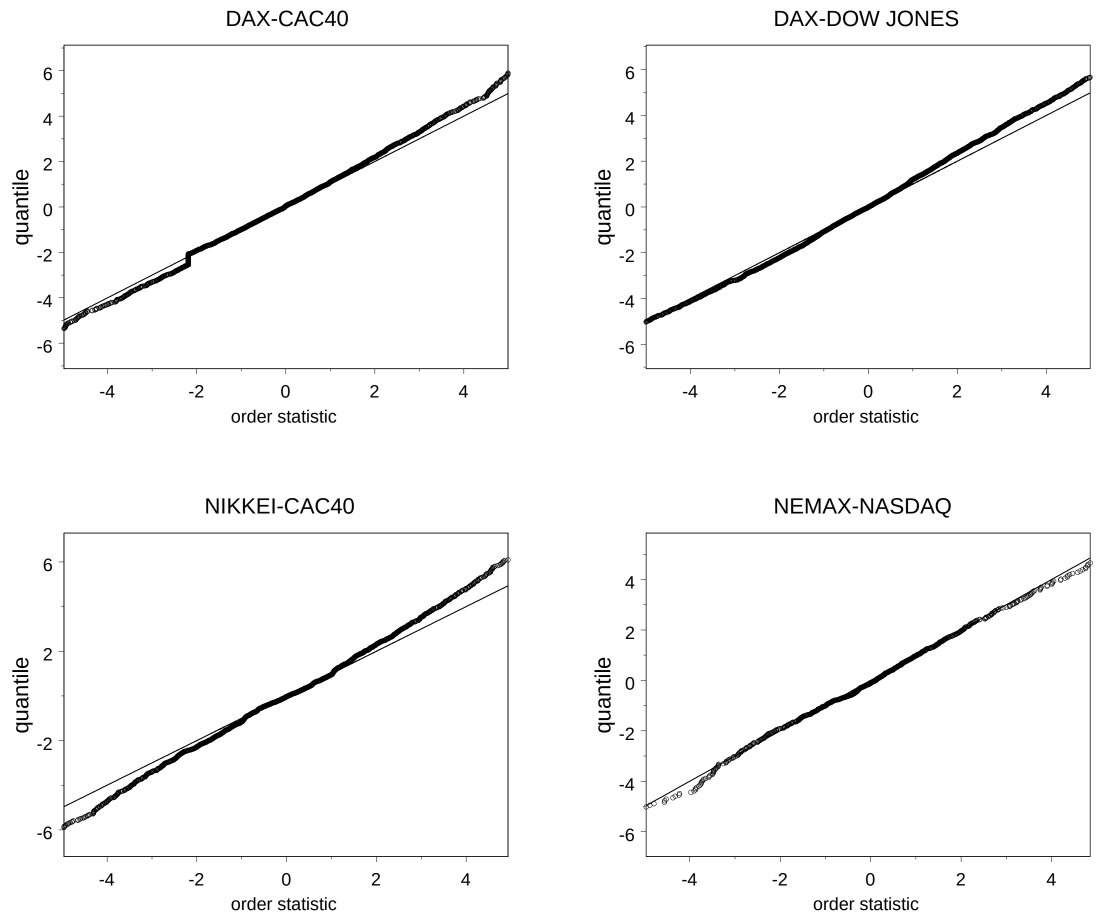


Figure 5.11: QQ-plots as introduced in Section 5.2.4 of bivariate log-returns of four asset series

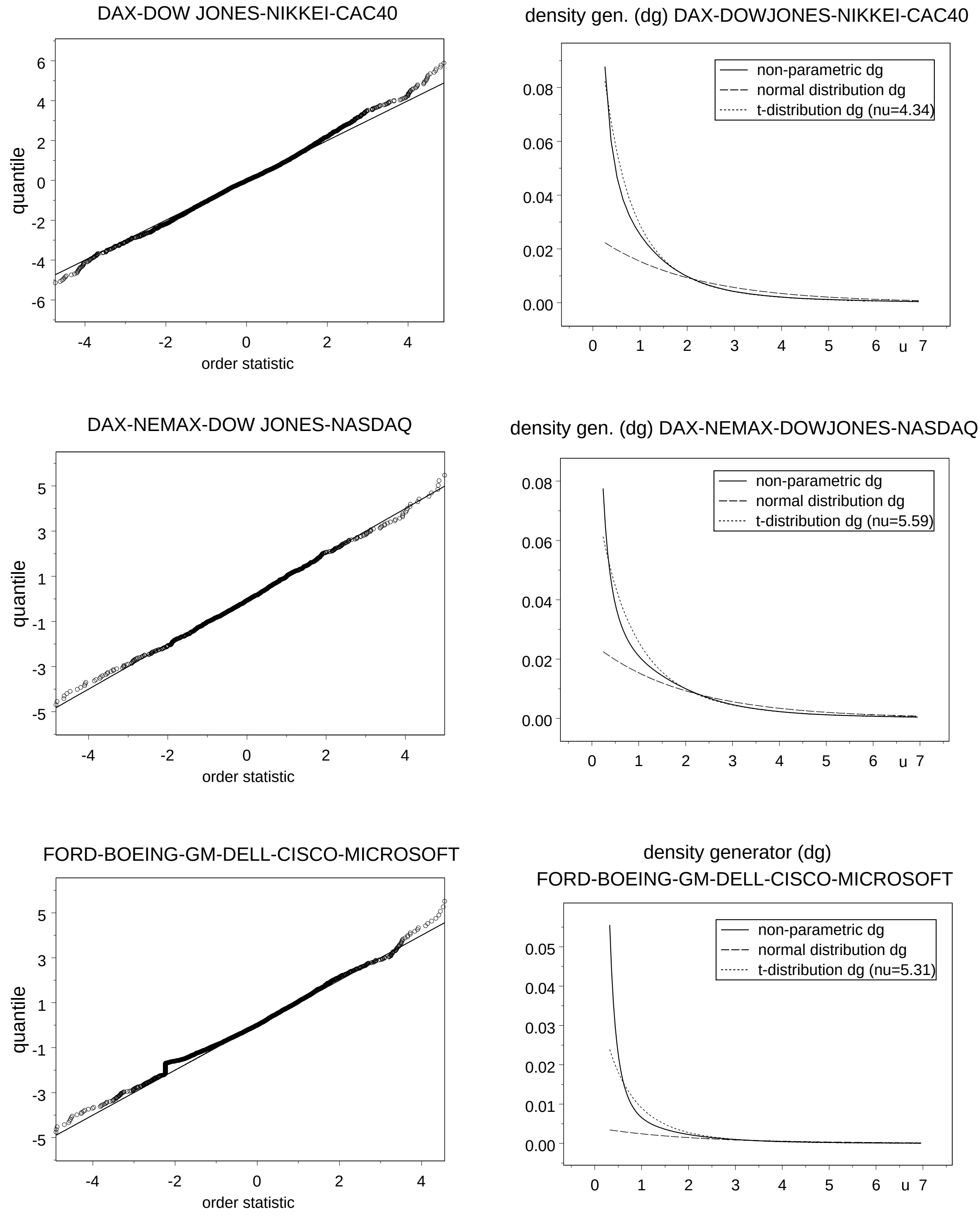


Figure 5.12: QQ-plots as introduced in Section 5.2.4 and density generators for multidimensional log-returns of DAX-DOWJONES-NIKKEI-CAC40 from 1987 until 2002 (first level), of DAX-NEMAX-DOWJONES-NASDAQ from 1998 until 2002 (second level) and of FORD-BOEING-GM-DELL-CISCO-MICROSOFT from 1990 until 2002 (third level).

Portfolio analysis

The Value-at-Risk of an asset portfolio with n assets and portfolio-value functions $p_t : \mathbb{R}^n \to \mathbb{R},\ t \in \mathbb{N}$, at some discrete time points $t = 0, \ldots, m,\ m \in \mathbb{N}$ is now defined by

$$VaR_t^\alpha = \mathrm{argsup}_{y \in \mathbb{R}} \mathbb{P}(\Delta p_t \leq y) \leq \alpha \tag{5.39}$$

to a confidence level $\alpha > 0$. Here the random vectors $X^{(t)},\ t = 0, \ldots, m,\ m \in \mathbb{N}$ contain the nominal asset values at time t and the absolute portfolio return is defined by $\Delta p_t := p_t - p_{t-1}$. We assume that $\Delta p_t(X^{(t)}, X^{(t-1)})$ depends only on the componentwise relative asset-returns at time t and the nominal asset values $X^{(t-1)}$. In particular, we define the random variable V_t^p describing the portfolio's absolute return at time t by

$$V_t^p := \Delta p_t(X^{(t)}, X^{(t-1)}) = f_t(\Delta_{rel} X_t, X^{(t-1)})$$

for some functions $f_t : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ and $\Delta_{rel} X_t := (\log(X_i^{(t)}) - \log(X_i^{(t-1)}))_{i=1,\ldots,n}$. For modelling reasons we suppose the log-returns $\Delta_{rel} X_t,\ t = 1, \ldots, m,\ m \in \mathbb{N}$ to be iid observations. The main intention of this section is now formulated by the assumption that $\Delta_{rel} X_t$ follows an elliptically contoured distribution:

$$\Delta_{rel} X_t \sim E_n(\mu, \Sigma, g). \tag{5.40}$$

For reasons of simplicity we do not utilize the innovations of an autoregressive-type process or an ARCH-type process, as in our context the results turned out to be quite similar.

In general, Value-at-Risk determination within modern portfolio theory utilizes one of four approaches: i) explicit VaR calculation, ii) structured Monte-Carlo simulation, iii) variance-covariance approximation, or iv) historical simulation. One of the main advantages of assumption (5.40) is that the portfolio's VaR can be explicitly calculated for every linear asset portfolio. Precisely, for a linear asset portfolio we obtain

$$V_t^p = X_{t-1}' \Delta_{rel} X_t.$$

The portfolio's VaR to a confidence level α, for a given realization x_{t-1} of the asset vector $X^{(t-1)}$ and under the assumption of elliptically contoured asset-returns $\Delta_{rel} X_t \in E_n(\mu, \Sigma, g)$ with positive-definite Σ is given by

$$VaR_t^\alpha = x_{t-1}' \mu - h(\alpha) \sqrt{x_{t-1}' \Sigma x_{t-1}}, \tag{5.41}$$

with $h(\alpha)$ being defined below. Formula (5.41) can be readily shown via the stochastic representation (3.16) of elliptically contoured asset-returns, i.e.,

$$\Delta_{rel} X_t \stackrel{d}{=} \mu + R_n A' U^{(n)}.$$

Then

$$V_t^p = x_{t-1}' \Delta_{rel} X_t \stackrel{d}{=} x_{t-1}' \mu + R_n (A x_{t-1})' U^{(n)}$$

is distributed according to an $E_1(x'_{t-1}\mu, x'_{t-1}\Sigma x_{t-1}, g)$ distribution. Define now $VaR_t^\alpha := x'_{t-1}\mu - h(\alpha)\sqrt{x'_{t-1}\Sigma x_{t-1}}$ with $h(\alpha)$ the $1-2\alpha$ quantile of the positive random variable $R_n B$, where B^2 is $Beta(1/2, (n-1)/2)$ distributed and independent of R_n. Observe that

$$\begin{aligned}\mathbb{P}(x'_{t-1}\Delta_{rel}X_t \leq VaR_t^\alpha) &= \mathbb{P}(x'_{t-1}\mu + R_n(Ax_{t-1})'U^{(n)} \leq VaR_t^\alpha)\\ &= \mathbb{P}(R_n BU^{(1)} \leq -h(\alpha)) = \frac{1}{2}\mathbb{P}(R_n B \geq h(\alpha)) = \alpha,\end{aligned}$$

where the last but one equation follows by Theorem 2.15 in Fang, Kotz, and Ng (1990b). Now the assertion of formula (5.41) is shown.

Tables 5.6 and 5.7 show various VaR figures for some asset portfolios we considered before. First, we present the empirical VaRs for linear portfolios with equally weighted assets to the confidence levels $\alpha = 0.01,\ 0.025,\ 0.05$. Further, we calculate the VaRs analytically via different density generators, namely via the nonparametric density generator and the density generators of the fitted t-distribution, the normal distribution and generalized hyperbolic distribution (GH). Finally, a simulation study shows the finite-sample properties of the corresponding VaR estimations.

Data analysis. The analytical VaR figures in Table 5.6 reveal that the normal distribution underestimates the 0.01 and 0.025 VaR's consequently whereas the 0.05 VaR is estimated reasonable. In contrast, the semi-parametric elliptically contoured distribution and the t-distribution perform better for the higher 0.01 and 0.025 quantiles than for the 0.05 VaR. The GH-distribution underestimates the VaR for nearly every asset combination. For the simulated VaR figures in Table 5.7 we obtain a similar picture. As a result we suggest the semi-parametric model as a suitable substitute for the above parametric elliptically contoured distributions in the context of portfolio VaR calculations.

Comparison of 'old' and 'new' economy. We would expect (as we know from the collapse of the dotcom bubble) that 'high-tech', 'new' economy stocks (indices NEMAX-NASDAQ) are more risky than traditional 'old' economy stocks (indices DAX-DOW JONES). We would thus expect fatter tails - slower decay of the density generator g - in the 'new' case then in the 'old'. Indeed, graphing the tails of the density generator (see Figure 5.10) confirms our expectations: 'new' lies above 'old' in the tails. Note however that the comparison is reversed in the bulk of data, where in particular the degrees of freedom (df) of the approximating t-distributions are determined. Thus 'new' has df 5.85 while 'old' has 5.35. Within the t-family, higher df means thinner tails; the crossing of the graphs of the density generator explains this reversal.

Note also the much bigger 1% VaR, -61.4 to -36.8, of 'new' against 'old'. Referring to (5.41), note that 'new' and 'old' have similar $h(\alpha)$; the difference is accounted for by the different covariance matrices Σ.

Table 5.6: Empirical and analytical VaR's to the confidence levels $\alpha = 0.01,\ 0.025,\ 0.05$ for various linear portfolios with equally weighted assets.

Assets	empirical VAR			semi-parametric VAR			normal VAR		
	$\alpha = 0.01$	0.025	0.05	$\alpha = 0.01$	0.025	0.05	$\alpha = 0.01$	0.025	0.05
DAX-DOWJONES	-36.76	-28.37	-22.19	-35.63	-26.75	-20.82	-31.41	-26.48	-22.24
NEMAX-NASDAQ	-61.41	-46.81	-38.28	-61.05	-47.27	-36.71	-53.85	-45.43	-38.20
DAX-CAC40	-38.47	-27.61	-19.37	-32.91	-24.96	-19.57	-29.66	-24.96	-20.92
NIKKEI-CAC40	-29.95	-22.86	-18.50	-29.14	-22.30	-17.23	-25.68	-21.64	-18.17
DAX-DOWJONES-NIKKEI-CAC40	-28.13	-19.83	-15.13	-23.53	-17.19	-12.19	-22.25	-18.73	-15.70
DAX-NEMAX-DOWJONES-NASDAQ	-44.05	-35.78	-28.31	-44.25	-32.32	-23.45	-39.93	-33.68	-28.31
FORD-BOEING-GM-DELL-CISCO-MICROSOFT	-42.65	-34.15	-27.69	-44.97	-32.34	-20.39	-38.93	-32.68	-27.30

Assets	empirical VAR			t VAR			GH VAR		
	$\alpha = 0.01$	0.025	0.05	$\alpha = 0.01$	0.025	0.05	$\alpha = 0.01$	0.025	0.05
DAX-DOWJONES	-36.76	-28.37	-22.19	-34.95	-26.93	-21.24	-32.87	-25.74	-20.14
NEMAX-NASDAQ	-61.41	-46.81	-38.28	-59.45	-46.28	-36.66	-58.33	-45.52	-35.79
DAX-CAC40	-38.47	-27.61	-19.37	-33.65	-25.18	-19.47	-31.47	-24.30	-18.79
NIKKEI-CAC40	-29.95	-22.86	-18.50	-29.19	-21.71	-16.68	-28.51	-21.92	-16.91
DAX-DOWJONES-NIKKEI-CAC40	-28.13	-19.83	-15.13	-25.23	-18.89	-14.62	-22.58	-17.52	-13.64
DAX-NEMAX-DOWJONES-NASDAQ	-44.05	-35.78	-28.31	-44.25	-34.31	-27.17	-42.10	-32.80	-25.37
FORD-BOEING-GM-DELL-CISCO-MICROSOFT	-42.65	-34.15	-27.69	-43.43	-33.26	-26.04	-47.64	-36.50	-28.23

Table 5.7: Mean (standard deviation) of 100 simulated VaR-figures based on 2000 equally weighted asset-return data.

Assets	semi-parametric VAR			normal VAR		
	$\alpha = 0.01$	0.025	0.05	$\alpha = 0.01$	0.025	0.05
DAX-DOWJONES	-35.93 (2.21)	-26.76 (1.17)	-21.05 (0.73)	-31.49 (1.16)	-26.44 (0.84)	-22.21 (0.63)
NEMAX-NASDAQ	-60.64 (3.41)	-47.32 (2.46)	-36.90 (1.47)	-54.09 (1.93)	-45.66 (1.40)	-38.27 (1.21)
DAX-CAC40	-33.85 (1.86)	-26.03 (1.33)	-20.28 (0.71)	-29.76 (1.14)	-24.96 (0.71)	-21.00 (0.57)
NIKKEI-CAC40	-29.24 (1.90)	-22.26 (1.08)	-17.18 (0.77)	-25.83 (0.90)	-21.64 (0.58)	-18.18 (0.55)
DAX-DOWJONES-NIKKEI-CAC40	-21.10 (0.97)	-16.95 (0.61)	-13.48 (0.46)	-22.19 (0.71)	-18.78 (0.50)	-15.74 (0.41)
DAX-NEMAX-DOWJONES-NASDAQ	-41.34 (1.74)	-33.25 (1.34)	-26.65 (1.02)	-40.04 (1.27)	-33.80 (0.92)	-28.41 (0.73)
FORD-BOEING-GM-DELL-CISCO-MICROSOFT	-40.28 (1.20)	-32.27 (1.22)	-25.89 (0.89)	-38.93 (1.43)	-32.57 (1.10)	-27.21 (0.92)

Assets	t VAR			GH VAR		
	$\alpha = 0.01$	0.025	0.05	$\alpha = 0.01$	0.025	0.05
DAX-DOWJONES	-35.31 (2.14)	-27.12 (1.33)	-21.37 (0.84)	-34.56 (1.64)	-27.12 (1.14)	-21.30 (0.73)
NEMAX-NASDAQ	-59.83 (2.90)	-46.39 (2.02)	-36.91 (1.37)	-57.27 (2.86)	-44.91 (1.99)	-35.56 (1.34)
DAX-CAC40	-33.74 (2.42)	-25.13 (1.26)	-19.43 (0.73)	-31.97 (1.56)	-24.76 (0.96)	-19.24 (0.67)
NIKKEI-CAC40	-29.73 (2.30)	-21.81 (1.13)	-16.72 (0.69)	-29.56 (1.73)	-22.71 (1.03)	-17.65 (0.72)
DAX-DOWJONES-NIKKEI-CAC40	-25.41 (1.84)	-18.89 (1.02)	-14.56 (0.62)	-22.88 (1.15)	-17.55 (0.69)	-13.74 (0.51)
DAX-NEMAX-DOWJONES-NASDAQ	-44.65 (2.44)	-34.35 (1.62)	-27.29 (1.06)	-43.30 (2.22)	-33.83 (1.42)	-26.49 (1.08)
FORD-BOEING-GM-DELL-CISCO-MICROSOFT	-43.97 (2.45)	-33.51 (1.54)	-26.22 (1.03)	-46.69 (2.77)	-35.73 (1.69)	-27.36 (1.25)

The problem addressed in this section is the modelling of stock-price and asset-return distributions in higher dimensions, motivated by questions of portfolio selection and risk management in finance. The model proposed here is semi-parametric, and uses elliptically contoured distributions, specifically normal variance mixtures with self-decomposable mixing distributions, with particular emphasis on the density generator g. In our view, this approach has much to recommend it: it has very nice theoretical properties, is easy and convenient to implement in simulation studies, and provides a good fit to a range of real financial data sets.

Many interesting theoretical and practical questions remain. We single out two: extension from normal variance mixtures to normal mean-variance mixtures (which leads beyond the elliptically contoured framework), and integration of our approach with the most closely related recent and contemporary work, that of Barndorff-Nielsen, Jensen, and Sørensen (1998), Korsholm (2000), Hodgson, Linton, and Vorkink (2000) and Barndorff-Nielsen and Shephard (2001).

Chapter 6

Special applications in finance

6.1 Risk-based pricing

The results of the present section have been established in the research project "Risk-based Pricing" during a collaboration between the Department of Number Theory and Probability Theory at the University of Ulm and the Department of Information Mining at DaimlerChrysler AG, Research and Technology, in Ulm. Further, the results are part of the project "Accord" at DaimlerChrysler Services in Berlin.[1]

The supply of credit capital is one of the fundamental tasks of financial institutions and services. Consequently, these institutions and services are exposed to *credit risk*, i.e., the risk of losses caused by the deterioration of the borrower's or contract's credit quality. Credit risk can be decomposed into default risk and migration risk. Default risk relates to possible counter-party defaults, whereas migration risk arises due to declining credit qualities. Most institutions which are strategically focusing on lending and leasing activities hold credit risk positions through maturity. As a consequence, they are substantially exposed to default risk rather than to migration risk. Therefore, we mainly focus on default risk in the remaining section. Traditionally, the pricing of credit contracts does not include appropriate or even any compensation (premium) for credit risk (nor market risk). The following elaboration is placed into that gap by developing a short methodological framework for *risk-based pricing* (RBP).

The main target of RBP relates to an increase of the profitability of credit lending businesses whereas profitability is measured via so-called *risk adjusted returns*. RBP methods are either driven by the market via *adverse selection*, by national regulators, or by the internal risk controlling. Adverse selection occurs if the credit prices are too small for contracts with low credibility and too large for contracts with high credibility. In that case the portfolio proportion of low-quality customers increases and the proportion of high-quality customers decreases, which deteriorates the portfolio's profitability and the institution's credit worthiness. Companies which have to fulfill the national

[1]Within the framework of the mentioned projects I would like to thank Professor Gholamreza Nakhaeizadeh, Professor Ulrich Stadtmüller, Angelika Oertel, Dr. Kai Bartlmae and the Risk Group team in Ulm.

regulatory requirements in the line of the New Basel Capital Accord are recommended to implement an RBP system if they are eligible for the Internal Ratings-based Approach (IRB) (see Basel Committee on Banking Supervision (2001) or Section 6.2 of the present thesis).

The following notions are necessary in the context of RBP. The *expected loss* is defined as the average loss anticipated for either a single credit contract or the entire credit portfolio due to credit risk within a certain time period. The expected loss of the portfolio equals the sum of the expected losses assigned to each contract in that portfolio. As the expected loss describes the average loss and does not cover the uncertainty of credit losses, we introduce two additional kinds of losses which are commonly referred to as the *unexpected loss* and the *extreme loss*. The unexpected loss is defined as the difference between a specified loss quantile of the portfolio' s or contract's loss distribution and the expected loss for a given time period (see also Figure 6.1). Extreme losses refer to large loss events which are often induced by dependencies between distinct credit-defaults.

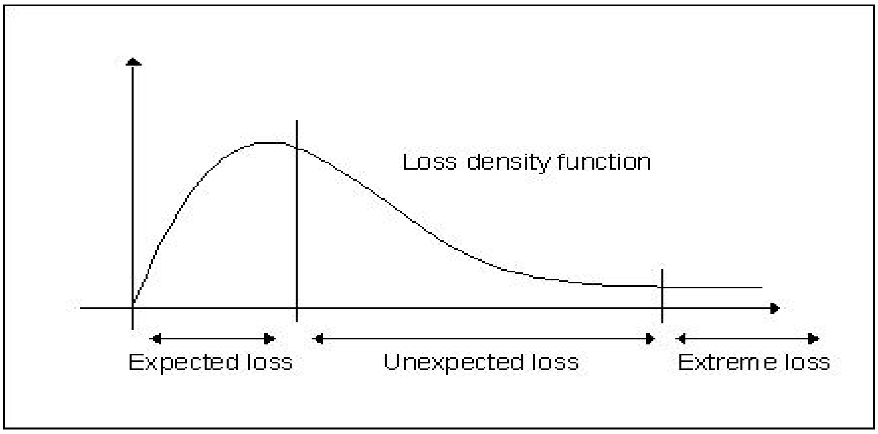


Figure 6.1: Expected, unexpected, and extreme loss.

It is imperative that each financial institution provides sufficient *risk capital* to sustain the above fluctuations in the credit portfolio value or credit contract value in order to protect the liquidity and financial stability of the financial institution. Risk capital mainly consists of equity or equity-like capital (so-called Tier 1,2, and 3 capital, cf. Ong (1999)). Often, the unexpected loss of the portfolio equals the internal supply of risk capital. Note that in contrast to the unexpected loss, the expected loss is considered as "cost of doing business".

The applicability of RBP can be divided into *progressive* and *retrograde pricing*. Progressive pricing assesses and quantifies the risk contribution of a *new* credit contract towards a portfolio of credit risky assets, where the risk contribution is measured via the expected and unexpected loss of the contract. RBP transfers a certain part of the quantified risk contribution of the new contract to the borrower via a risk adjusted credit price or premium. On the other hand, retrograde pricing measures the risk contribution of each credit contract within a given, already existing, portfolio. Active risk oriented portfolio management requires retrograde pricing for its decision making processes. Besides the detection of potential profitability gaps, retrograde pricing reveals portfolio risk concentrations and improves the portfolio's overall risk-return ratio.

RBP models Several RBP methodologies have been developed in the last decade. Most of them can be distinguished by using either market data, like time series of stock prices or bond spreads, or exploiting internal credit-risk data, like borrower ratings or scorings. The first kind of methodology usually applies the well-known theoretical framework of option pricing in the sense of Merton, Black, and Scholes. However, this methodology requires the counter parties to be listed on the stock or bond market which is obviously not fulfilled by most retail businesses, small, and medium companies. In this thesis we concentrate on the second kind of model which uses internal credit risk data for the pricing.

All RBP methodologies require a *portfolio model* which describes the portfolio's risk structure in order to quantify the credit risk. However, for example, the calculation of risk capital depends heavily on the choice of the portfolio model. The portfolio model we apply in this section incorporates the recent standards of credit risk modelling, and takes into account the current data limitation of credit risk parameters. Precisely, a rating/scoring-based "risk-bucketing" system determines the contract's credit risk contribution and consequently the related costs for credit risk (see also Gordy (2001)). In other words, all contracts are grouped into risk-homogeneous buckets for which a fixed risk-based costing rate per contract is calculated.

The price components The development of the RBP model follows a top down approach. In a first step, the risk adjusted total price is decomposed into three price components which will be explained and described below. In a second step, the mathematical pricing methodologies are developed for each component and the necessary input parameters are determined.

The price of a credit contract within the RBP framework consists of three price components:

1. the price premium for *administrative costs* (C^{Ad}),
2. the price premium for *costs for expected loss* (C^{EL}), and
3. the price premium for *funding costs* (C^{F}).

Administrative costs arise within the origination process of the credit contract (e.g. credit application, evaluation, and approval costs) and the following administration process. These costs are assumed to be risk neutral. Administrative costs arising due to a credit default are included in the so-called *loss given default* quantity which is defined below. The expected loss originates costs due to the average amount of credit loss and is assumed to be certain in the long run. Thus, the expected loss does not represent a risk position in itself and is seen as a predictable expense which influences the earnings or returns of credit lending. Consequently, the costs of expected loss are completely transferred towards the premium or price of the contract. Funding costs are decomposable into *costs for refinancing* and *costs for risk capital supply.* The costs for risk capital supply are determined by multiplying the institute-specific interest rate on risk capital (so-called *hurdle rate*) with the provided amount of risk capital.

Summarizing the above, RBP provides the *(minimum price)* or *(minimum) interest rate* r of a credit contract within a credit risk coherent environment under a given target of risk adjusted return.

Portfolio model Credit portfolio models provide as primary output the distribution of the portfolio's credit losses, help to identify concentrations of credit risk in the portfolio, and reveal opportunities for risk diversification. In our framework, the expected loss plus the unexpected loss equals a high quantile of the portfolio's loss distribution under a certain confidence level for a specified time horizon. This is closely related to the concept of Value-at-Risk (see also Section 5.2.4). The following model comprises the portfolio model utilized in the Internal Ratings-based Approach of the New Basel Capital Accord (cf. Basel Committee on Banking Supervision (2001)) as a special case. Recall, the final aim is to group all credit assets into risk-segments which are presumed to be homogenous regarding the relevant risk drivers. The most important risk drivers are:

1. The *probability of default* (PD), which is defined as the likelihood that a borrower enters a specified default status within a particular time period.
2. The *loss given default* (LGD), which is a measure of the expected average loss experienced per unit of exposure if its counter-party defaults.
3. The *exposure at default* (EAD), which is defined as the amount to which the lender is exposed to the borrower at the time of default.

For reasons of simplicity we assume that each borrower is uniquely assigned to one contract.

The portfolio's *loss distribution* is defined via a random variable $L_{(n)}$ which describes the portfolio loss, given n contracts (or borrowers), for a particular evaluation horizon, i.e.,

$$L_{(n)} = \sum_{i=1}^{n} L_i := \sum_{i=1}^{n} D_i \cdot LGD_i \cdot EAD_i, \tag{6.1}$$

where the Bernoulli random variable D_i defines the *default status* of the borrower (1=default, 0=nondefault), LGD_i denotes the loss given default, and EAD_i denotes the exposure at default of contract $i = 1, \ldots, n$. The default status D_i is modelled via a latent random variable X_i representing the returns on the borrower's assets, i.e.,

$$D_i \stackrel{d}{=} \mathbf{1}_{\{X_i \leq \gamma_i\}},$$

where γ_i is referred to as default barrier. Therefore, the borrower defaults if the asset return falls below a pre-specified default barrier.

Regarding the modelling of X_i, a classical *one-factor model* is applied which incorporates several advantages and which reflects the current credit lending practices (see "Credit risk modelling: Current practices and applications", Basel Committee on Banking Supervision (1999)). Although the one-factor model is not a flexible portfolio model, it is sophisticated enough to incorporate the limited internal risk parameter information available in the near or medium term. Moreover, according to Gordy (2001), markets do not provide precise information on correlation of credit events across customers yet and the empirical challenge would likely concern the quality of default probabilities estimates. Especially for portfolios of retail and small business, the latter portfolio model seems to be appropriate.

The random variable X_i which describes the returns of the borrower's assets is distributed according to

$$X_i \stackrel{d}{=} \omega_i Z + \xi_i \varepsilon_i, \quad i = 1, \ldots, n, \tag{6.2}$$

where Z denotes the common *systematic risk factor* related to all n borrowers and $\varepsilon_i, \quad i = 1, \ldots, n$, relates to the *borrower specific risk*. The random variable Z and the random vector ε are uncorrelated and spherically distributed (see Definition 3.1.1). The constants ω_i and ξ_i are referred to as *factor loadings* and determine the dependence towards the systematic risk factor Z. Thus, $\omega_i, \ i = 1, \ldots, n$, determine the overall dependence structure of the asset return vector $X = (X_1, \ldots, X_n)'$.

Consequently, the probability of default PD_i corresponding to borrower i in a certain time period is derived in the following way:

$$PD_i = \mathbb{P}(D_i = 1) = \mathbb{P}(X_i \le \gamma_i) = \mathbb{P}(\omega_i Z + \xi_i \varepsilon_i \le \gamma_i). \tag{6.3}$$

Observe that equation (6.3) can be rewritten as

$$PD_i = \mathbb{P}\left(\frac{\omega_i}{\sqrt{\omega_i^2 + \xi_i^2}} Z + \frac{\xi_i}{\sqrt{\omega_i^2 + \xi_i^2}} \varepsilon_i \le \frac{\gamma_i}{\sqrt{\omega_i^2 + \xi_i^2}} \right).$$

Hence, the distribution of the components X_i of the asset return vector X are given by

$$X_i \stackrel{d}{=} \omega_i Z + \sqrt{1 - \omega_i^2}\, \varepsilon_i. \tag{6.4}$$

The corresponding matrix representation yields

$$X \stackrel{d}{=} \begin{pmatrix} \omega_1 & \sqrt{1-\omega_1^2} & 0 & \ldots & 0 \\ \omega_2 & 0 & \sqrt{1-\omega_2^2} & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ \omega_n & 0 & \ldots & 0 & \sqrt{1-\omega_n^2} \end{pmatrix} \cdot \begin{pmatrix} Z \\ \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

According to the findings in Chapters 2, 3, and 4, multivariate asset return distributions possessing the tail dependence property are crucial for credit portfolio modelling, since they are able to incorporate dependencies of extremal credit default events. Therefore, we propose to model the random vector $(Z, \varepsilon_1, \ldots, \varepsilon_n)'$ via some spherically contoured random vector possessing the tail-dependence property (cf. Theorem 3.1.4 in Section 3.1 of the present thesis). Especially the t-distribution (cf. Section 3.2.3) represents an appealing candidate. According to Section 3.2.1, the asset return random vector X is elliptically distributed and, according to Theorem 3.2.2 in Section 3.2.2, it runs out that X inherits tail dependent margins.

Finally, we provide more inside into the calculation of the costs arising due to expected loss and supply of risk capital.

Costs of expected loss The portfolio's expected loss is additive, i.e., the contract-specific expected losses add up to the total portfolio's expected loss. Therefore, we restrict ourselves to the calculation of the expected loss which is allocated to a single contract $i = 1, \ldots, n$. The expected loss EL_i of a single contract i equals the discounted sum of the expected losses EL_{ij} per payment period $j = 1, \ldots, m$ over the entire contractual lifetime, i.e.,

$$EL_i = \sum_{j=1}^{m} EL_{ij} \cdot d_j, \quad i = 1, \ldots, n,$$

where d_j denotes an appropriate discounting factor. Further, the expected loss EL_{ij} for contract i in a particular payment period j is calculated as the product of the exposure at default, the contract specific loss given default, and the probability of default. For simplicity we assume that the contract's payment structure and the loss given default are purely deterministic (otherwise we assume independence and take the expectation). Thus,

$$EL_i = \sum_{j=1}^{m} EL_{ij} \cdot d_j = \sum_{j=1}^{m} PD_{ij} \cdot LGD_{ij} \cdot EAD_{ij} \cdot d_j, \tag{6.5}$$

The costs of expected loss coincide with the amount of expected loss and are completely transferred to the borrower. Note that the present pricing concept utilizes the discounted cash flow structures of each credit contract.

Costs for risk capital supply Recall that the contract's funding costs are divided into the costs for refinancing and the costs for risk capital supply. The latter costs are determined by the contract's risk contribution to the overall portfolio's loss distribution. As we have mentioned before, credit risk managers must allocate risk capital on the portfolio level which equals the portfolio's unexpected loss. In particular, the institution assigns risk capital in order to obtain a target survival probability or target rating. In our context, the unexpected loss plus the expected loss corresponds to a high target quantile of the loss distribution given in equation (6.1), i.e., the VaR of the loss distribution. The VaR contribution of a single contract (marginal VaR or MVaR) to the total portfolio's VaR per payment period is determined via Monte Carlo simulation. Under further assumptions on the distribution of the asset return vector X, we may find an analytical formula for the MVaR (cf. Gordy (2001)).

Finally, the (minimum) risk-based price (or interest rate) of a credit contract is calculated via the following *iteration algorithm*:

1. Set the initial interest rate equal to zero e.g. $r = 0\%$.

2. Calculate the contractual cash-flow corresponding to r e.g. periodical outstanding, interest payment, installment, etc..
3. Determine the periodical risk parameters e.g. PD_i, LGD_i, EAD_i, etc..
4. Derive the contract specific expected loss over the entire contractual lifetime EL_i.
5. Derive the contract specific marginal VaR over the entire contractual lifetime.
6. Calculate the corresponding costs C^{Ad}, C^{EL}, and C^F.
7. If the costs C^{Ad}, C^{EL}, and C^F are **not** covered by the contractual interest payments: Increase r dynamically and goto step 1.
8. If the costs C^{Ad}, C^{EL}, and C^F are covered by the contractual interest payments: r and the corresponding (minimum) risk-based price are found.

6.2 The New Basel Capital Accord

The 1988 Basel Capital Accord is a current benchmark for many national regulatory laws related to "economic capital" on lending businesses of commercial banks. This Accord requires banks to keep an 8% capital charge of the loan face value for any commercial loan in order to cushion losses from an eventual credit default. An increasing problem arises from the overall 8% capital charge which does not include the financial strength of the borrower and the value of the collateral. This led to an off-balance-sheet movement of low-risk credits and a retainment of high-risk credits. To overcome the insufficiency of credit-risk differentiation and other inadequacies, the Basel Committee launched the New Basel Capital Accord - Basel II which is expected to become national law in 2006. Basel II is divided into the Standard approach and the Internal Ratings-Based approach (IRB). Despite the Standard approach, which basically reflects the 1988 Basel Accord, the IRB approach calculates the "economic capital" by using a credit risk portfolio model. Recall, by credit-risk portfolio model we understand a function which maps a set of instrument-level and market-level parameters (cf. Gordy (2001), p. 1) to a distribution for portfolio credit losses over a specified horizon. In this context "economic capital" denotes the VaR of the portfolio's loss distribution. Following the IRB approach banks are now required to derive probabilities of default per loan (or exposure) via internal estimations or external mappings. Together with other parameters the required regulatory capital increases with increasing probability of default and decreasing credit quality, respectively. Furthermore banks have an incentive to raise the "economic capital" in order to improve their own credit rating.

The IRB approach utilizes a single-factor model as portfolio model, as explained in Section 6.1, which describes credit defaults by a two-state Merton model. This model can be compared to a simplified framework of CreditMetrics (1998). In particular, borrower i is linked to a random variable X_i which represents the normalized return of its assets, i.e.,

$$X_i = \omega_i Z + \sqrt{1-\omega_i^2}\varepsilon_i, \qquad i = 1, \dots, n, \tag{6.6}$$

where Z is a single common systematic risk factor related to all n borrowers and $\varepsilon_i,\ i = 1, \ldots, n$, denotes the borrower-specific risk. The random variables Z and $\varepsilon_i,\ i = 1, \ldots, n$, are assumed to be standard normal distributed and mutually independent. The parameters $\omega_i,\ i = 1, \ldots, n$, are again called factor loadings, and regulate the sensitivity towards the systematic risk factor Z.

The simplicity of the above single factor model has a significant advantage: It provides portfolio-invariant capital charges, i.e., the charges depend only on the loan's own properties and not on the corresponding portfolio properties. According to Gordy and Heitfield (2000), p. 5, this is essential for an IRB capital regime.

Observe that the IRB credit risk portfolio model (6.6) makes use of a multivariate normal distribution, and thus the dependence structure of the portfolio's asset returns is that of a multivariate normal distribution. However, according to the findings of the present thesis, the normal distribution should be considered with care because of its inability to model dependence of extreme events or tail dependence.

Like in Section 6.1, we propose to substitute the dependence structure of the above multivariate normal distribution by the dependence structure of elliptically contoured distributions which possess the tail dependence property (cf. Theorem 3.2.2 in Section 3.2.2). The class of elliptically contoured distributions inherits many of the properties established in the IRB credit risk portfolio model (cf. Gordy (2001)).

6.3 Project results

The present section gives a short overview of the results which have been elaborated within the project "Risk-based Pricing" during the collaboration between the Department of Number Theory and Probability Theory at the University of Ulm and the Department of Information Mining at DaimlerChrysler AG, Research and Technology, in Ulm.

The insufficiency of multivariate normal distributions to model dependencies of extreme events (denoted by tail dependence), like dependencies of large credit loss events, has indeed been found in many financial data during the project phase. According to the theoretical findings and simulation studies of the present thesis, the family of elliptically contoured distributions possessing tail dependence turns out to be appropriate to substitute the multivariate normal distributions in the one-factor portfolio models which have been considered in Sections 6.1 and 6.2. From the practical point of view, this kind of substitution seems to be new. Several empirical studies of portfolio data within the project reveal that the latter family of distributions is well suited for application in risk management. The model properties and results we obtained turn out to be satisfying for a sophisticated risk management environment. The mathematical foundation and exploration of extremal dependence or tail dependence within the class of elliptically contoured distributions was established in Chapters 3 and 4 of the present thesis.

In particular, the RBP methodology in Section 6.1 has been successfully tested and implemented within the project "Risk-based Pricing". The algorithms and results in

Sections 6.1 and 6.2 serve as primary input for several risk-management guidelines and for a prototypic software implementation. The software implementation provides an integrative pricing tool for lease and finance contracts, and utilizes the above RBP methodology. For the development, calibration and application of the RBP tool, the relevance of various risk parameters was brought up in Section 6.1. Along these lines we have elaborated methodologies to estimate the risk parameters from internal portfolio data; among them we mention the estimation of the probability of default. Besides, we developed a framework for backtesting the latter estimates of risk parameters.

6.4 The dynamics of tail dependence: Another empirical survey

The present section refers to Chapter 4. In particular, we provide further empirical surveys to illustrate the behavior and properties of the parametric and nonparametric estimates $\hat{\lambda}_U^E$ (see formula (4.39)) and $\hat{\lambda}_U$ (see formula (4.49)) for the tail-dependence coefficient introduced in Sections 4.9.1 and 4.9.2. Recall, the first type of estimate was specifically developed for the family of elliptically contoured distributions.

The figures given below emphasize that the parametric and nonparametric TDC estimates are applicable for various kinds of asset-return data. Beside stock and index returns, we also consider returns of exchange rates and bonds. Therefore, valuable information on dependencies of extreme events can be drawn for decisions within asset portfolio management.

All figures imply that the increase and decrease of tail dependence over time often behaves similar to the increase and decrease of the corresponding correlation. However, the magnitude between tail dependence and correlation might be very different. For example, Figure 6.2 reveals that the tail-dependence coefficients between the log-return of DAX and DOW JONES were similarly high in 1992 and 2002, whereas the corresponding correlation coefficients showed a quite different behavior. Figure 6.3 shows similar effects. Figure 6.4 reveals that the parametric TDC estimations perform quite well if tail dependence is very low. Further, all figures show that nonparametric TDC estimates are reasonable alternatives to parametric TDC estimates in case the underlying distributions are elliptically contoured.

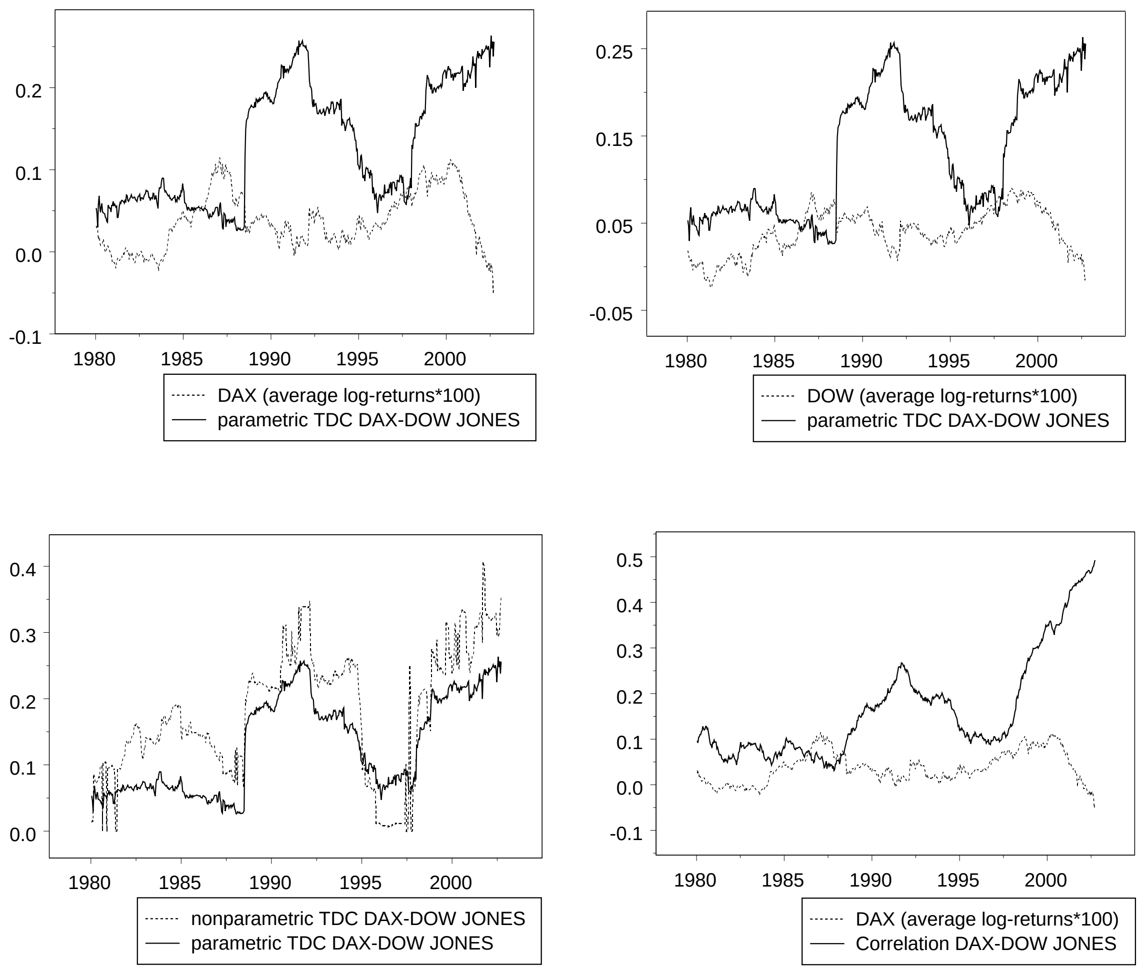


Figure 6.2: Data set: DAX and DOW JONES daily index log-returns. Upper plots: Parametric (lower) tail-dependence coefficient (TDC) estimates given in (4.39) over time. Lower left plot: Comparison between parametric and nonparametric (lower) TDC estimates given in (4.39) and (4.49), respectively. Lower right plot: Empirical correlation coefficients over time. All estimates are calculated for the years 1975-2002 with a horizon of 1000 trading days and a successive delay of 11 trading days.

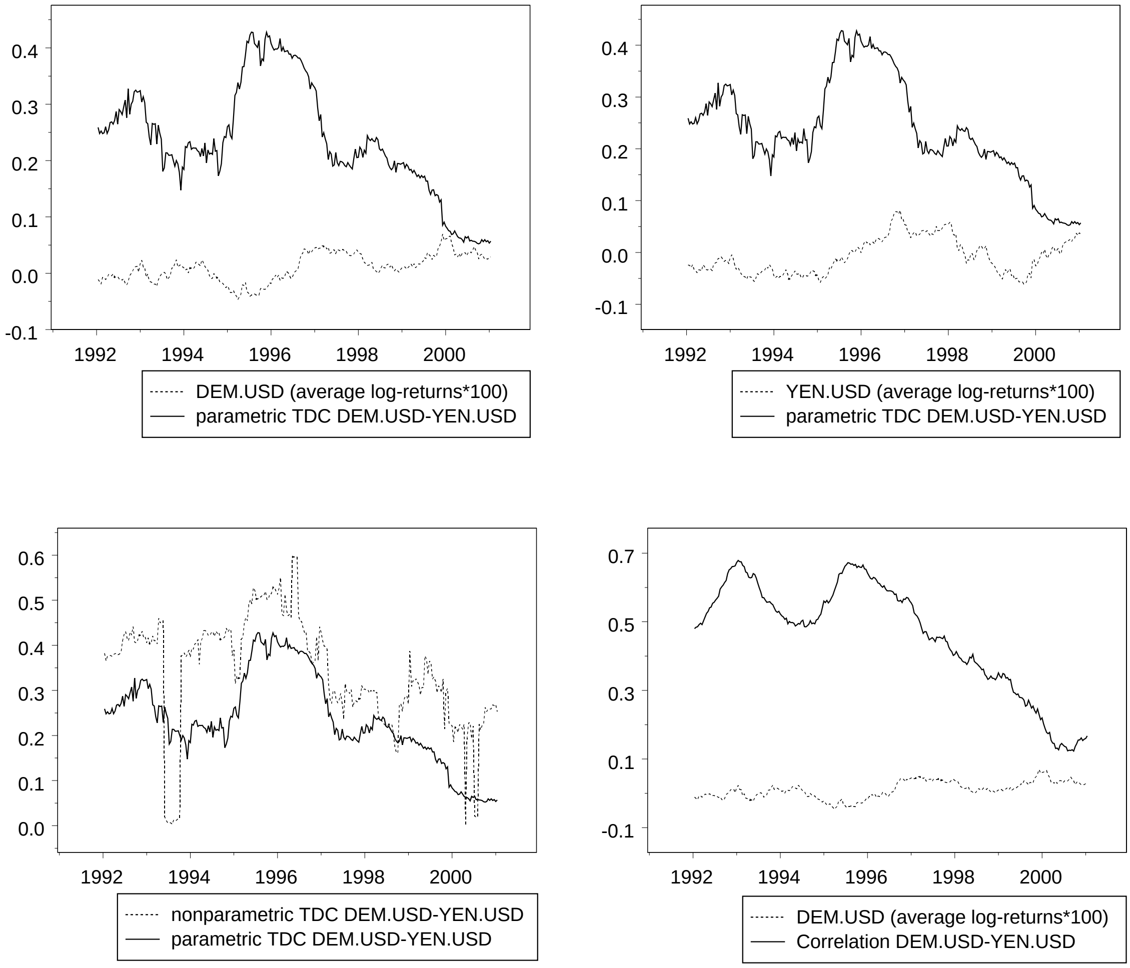


Figure 6.3: Data set: DM.USD and YEN.USD exchange rate log-returns. Upper plots: Parametric (lower) tail-dependence coefficient (TDC) estimates given in (4.39) over time. Lower left plot: Comparison between parametric and nonparametric (lower) TDC estimates given in (4.39) and (4.49), respectively. Lower right plot: Empirical correlation coefficients over time. All estimates are calculated for the years 1989-2002 with a horizon of 500 trading days and a successive delay of 10 trading days.

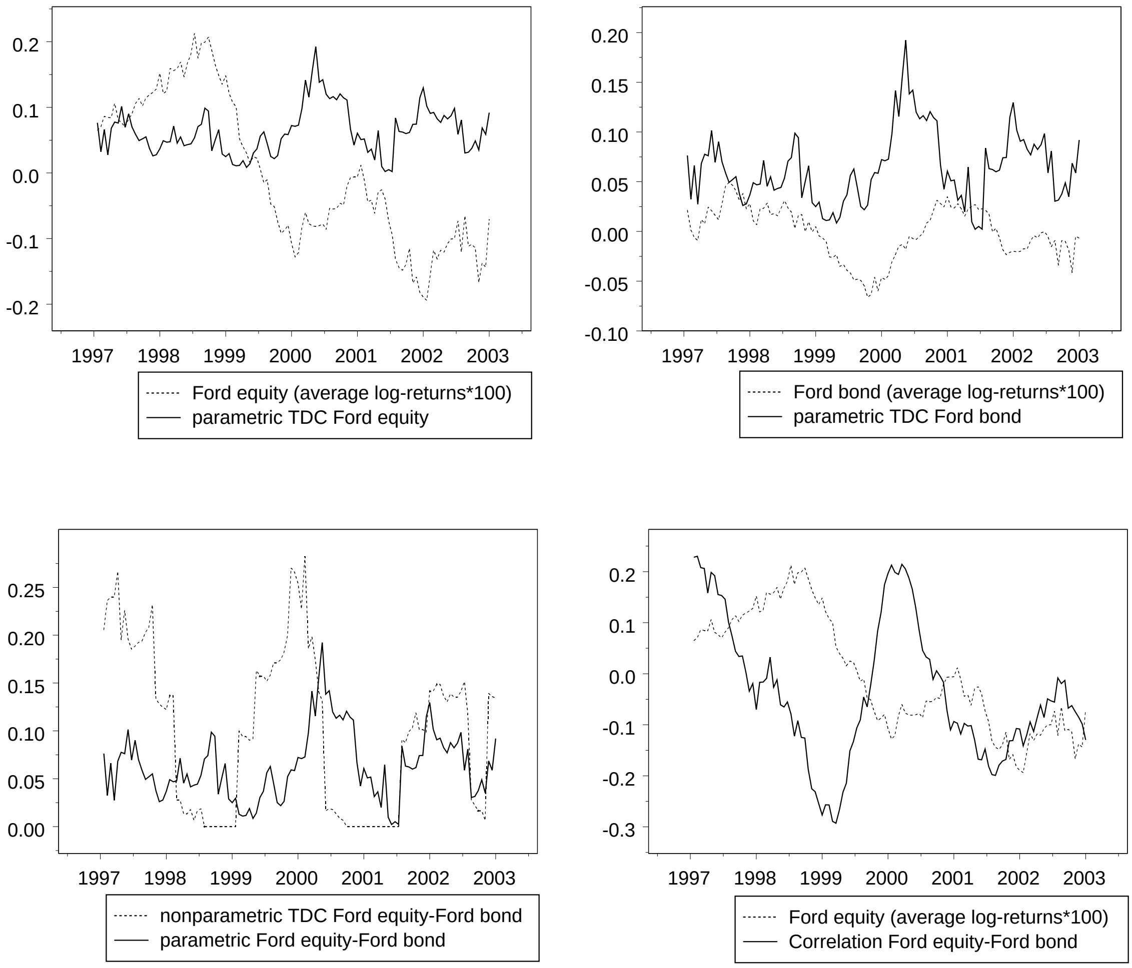


Figure 6.4: Data set: Ford Equity and Ford Bond log-returns. Upper plots: Parametric (lower) tail-dependence coefficient (TDC) estimates given in (4.39) over time. Lower left plot: Comparison between parametric and nonparametric (lower) TDC estimates given in (4.39) and (4.49), respectively. Lower right plot: Empirical correlation coefficients over time. All estimates are calculated for the years 1996-2002 with a horizon of 200 trading days and a successive delay of 10 trading days.

Appendix A

Supplementary

The proof of Theorem 4.5.3 (asymptotic normality of the tail-copula estimator) in Section 4.5 needs the following lemma. The lemma is stated in the original version as provided in Van der Vaart and Wellner (1996), p. 388. The appropriate space and the corresponding metric have to be adjusted.

Let $\mathcal{Y}$ and $\mathcal{Z}$ be subsets of normed spaces. Consider the maps $A : \mathcal{X} \mapsto \mathcal{Y}$ and $B : \mathcal{Y} \mapsto \mathcal{Z}$ which define the composition map $\phi(A, B) : \mathcal{X} \mapsto \mathcal{Z}$, by

$$\phi(A, B)(x) = B \circ A(x) = B(A(x)).$$

If B is a uniformly norm-bounded map from $\mathcal{Y} \mapsto \mathcal{Z}$, then $\phi(A, B)$ is a uniformly norm-bounded map from $\mathcal{X} \mapsto \mathcal{Z}$. We consider now ϕ as a map with domain $l^\infty(\mathcal{X}, \mathcal{Y}) \times l^\infty(\mathcal{Y}, \mathcal{Z})$ equipped with the norm $||(A, B)||_\infty = \sup_x ||A(x)||_{\mathcal{Y}} \vee \sup_y ||B(y)||_{\mathcal{Z}}$.

Lemma A.0.1 *Suppose that $B : \mathcal{Y} \mapsto \mathcal{Z}$ is Fréchet-differentiable uniformly in y in the range of A with derivatives B'_y such that $y \mapsto B'_y$ is uniformly norm-bounded. Then the composition map $\phi : l^\infty(\mathcal{X}, \mathcal{Y}) \times l^\infty(\mathcal{Y}, \mathcal{Z}) \mapsto l^\infty(\mathcal{X}, \mathcal{Z})$ is Hadamard-differentiable at (A, B) tangentially to the set $l^\infty(\mathcal{X}, \mathcal{Z}) \times UC(\mathcal{Y}, \mathcal{Z})$. The derivative is given by*

$$\phi'_{A,B}(\alpha, \beta)(x) = \beta \circ A(x) + B'_{A(x)}\big(\alpha(x)\big), \quad x \in \mathcal{X}.$$

Bibliography

Abramowitz, M., and I. A. Stegun, 1964, *Handbook of Mathematical Functions.* (National Bureau of Standards US Department of Commerce).

Anderson, T.W., K.-T. Fang, and H. Hsu, 1986, Maximum-likelihood estimates and likelihood-ratio criteria for multivariate elliptically contoured distributions, *Canadian Journal of Statistics* 14, 55–59.

Ané, T., and C. Kharoubi, 2001, Dependence Structure and Risk Measure, *forthcoming in Journal of Business.*

Araujo, A., and E. Giné, 1980, *The Central Limit Theorem for Real and Banach Valued Random Variables.* (John Wiley & Sons).

Atkinson, A.C., 1982, The simulation of generalized inverse Gaussian and hyperbolic random variables, *SIAM Journal on Scientific Computing* 3, 502–515.

Balkema, A.A., and L. de Haan, 1974, Residual lifetime at great age, *Annals of Probability* 2, 792–804.

Baringhaus, L., 1991, Testing for spherical symmetry of a multivariate distribution, *Annals of Statistics* 19, 899–917.

Barndorff-Nielsen, O.E., J.L Jensen, and M. Sørensen, 1998, Some stationary processes in discrete and continuous time, *Advances in Applied Probability* 30, 989 – 1007.

Barndorff-Nielsen, O.E., J.T. Kent, and M. Sørensen, 1982a, Normal variance-mean mixtures and z-distributions, *International Statistical Review* 50, 145–159.

Barndorff-Nielsen, O.E., and N. Shephard, 2001, Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics, *Journal of the Royal Statistical Society B* 63, 167–241.

Barndorff-Nielsen, O. E., 1977, Exponentially decreasing distributions for the logarithm of particle size, *Proceedings of the Royal Society of London* A, 401–419.

Barndorff-Nielsen, O. E., 1978, Hyperbolic distributions and distributions on hyperbolae, *Scandinavian Journal of Statistics* pp. 151–157.

Barndorff-Nielsen, O. E., and P. Blæsild, 1981, *In: Hyperbolic distributions and ramifications: Contributions to theory and application,* eds.: C. Taillie, G. Patil, and B. Baldessari Dordrecht: Reidel vol. 4 of *Statistical Distributions in Scientific Work* pp. 19–44.

Barndorff-Nielsen, O. E., J. Kent, and M. Sørensen, 1982b, Normal Variance-Mean Mixtures and z-Distributions, *International Statistical Review* 50, 145–159.

Baron, S., E. Liflyand, and U. Stadtmüller, 2000, Complementary Spaces and Multipliers of Double Fourier Series for Functions of Bounded Variation, *Journal of Mathematical Analysis and Applications* 250, 706–721.

Basel Committee on Banking Supervision, 1999, Credit risk modelling: Current practices and applications, *Bank of International Settlement.*

Basel Committee on Banking Supervision, 2001, The Internal Ratings-Based Approach, *Bank of International Settlement.*

Beran, R., 1979, Testing for ellipsoidal symmetry of a multivariate density, *Annals of Statistics* 7, 150–162.

Bickel, P.J., C.A.J. Klaasen, Y. Ritov, and J.A. Wellner, 1998, *Efficient and adaptive estimation for semi-parametric models.* (Springer, New York) 1st ed. 1993, Johns Hopkins UP.

Billingsley, P., 1968, *Convergence of Probability Measures.* (John Wiley & Sons, New York).

Bingham, N.H., and R. Kiesel, 2001a, Disordered and complex systems, *In: Hyperbolic and semiparametric models in finance*, eds.: Sollich, P. and Coolen, A.C.C. and Hughston, L.P. and Streater, R.F., American Institute of Physics, pp. 275–288.

Bingham, N. H., C. M. Goldie, and J. L. Teugels, 1987, *Regular Variation* vol. 27 of *Encyclopedia of Mathematics and Its Applications.* (Cambridge University Press Cambridge).

Bingham, N. H., and A. Inoue, 2000, Tauberian and Mercerian theorems for systems of kernels, *Journal of Mathematical Analysis and Applications* pp. 177–197.

Bingham, N. H., and R. Kiesel, 2001b, Modelling asset returns with hyperbolic distributions, *In: Asset return distributions*, eds.: J. Knight and S. Satchell, Butterworth-Heinemann, pp. 1–20.

Bingham, N. H., and R. Kiesel, 2002, Semi-parametric modelling in finance: theoretical foundation, *Quantitative Finance 2* pp. 241–250.

Bingham, N. H., R. Kiesel, and R. Schmidt, 2002, Semi-parametric models in finance: Econometric applications, *submitted, www.mathematik.uni-ulm.de/finmath/.*

Blæsild, P., and J. L. Jensen, 1981, Multivariate distributions of hyperbolic type, *Statistical distributions in scientific work 4,* eds.: C. Taillie, G. Patil, and B. Baldessari, Dordrecht: Reidel, 45–66.

Boender, C.G.E., and A.H.G. Rinnooy Kan, 1987, Bayesian stopping rules for multistart global optimization methods, *Mathematical Programming* 37, 59–80.

Bouyé, E., V. Durrleman, A. Nikeghbali, G. Riboulet, and T. Roncalli, 2000, Copulas for finance a reading guide and some applications, *Groupe de Recherche Opérationnelle, Crédit Lyonnais, preprint.*

Büning, H., and G. Trenkler, 1978, *Nichtparametrische statistische Methoden.* (Walter de Gruyter Verlag Berlin, New York).

Cambanis, S., S. Huang, and G. Simons, 1981, On the Theory of Elliptically Contoured Distributions, *Journal of Multivariate Analysis* 11, 368–85.

Campbell, R., K. Koedijk, and P. Kofman, 2002, Increased Correlation in Bear Markets, *Financial Analysts Journal* Jan-Feb, 87–94.

Carr, P., E. Chang, and D.B. Madan, 1998, The Variance-Gamma process and option pricing, *European Finance Review* 2, 79–105.

Casella, G., and R. L. Berger, 2002, *Statistical Inference.* (Duxbury Press Pacific Grove) 2. edn.

Coles, S., J. Heffernan, and J. Tawn, 1999, Dependence measures for extreme value analysis, *Extremes* 2, 339–366.

CreditMetrics, 1998, Technical document, *JP. Morgan, New York.*

Cuculescu, I., and R. Theodorescu, 2001, Copulas: diagonals, tracks, *Rev. Roumaine Math. Pures Appl.* 46, 731–742.

Cuculescu, I., and R. Theodorescu, 2002, Unimodal copulas, *Universitatea Bucureşti and Université Laval, preprint.*

Dacorogna, M.M., U.A. Müller, and O.V. Pictet, 1998, Heavy tails in high-frequency financial data, *In: A practical guide to heavy tails. Statistical techniques and applications,* eds.: Adler, R.J. and Feldmann, R.E. and Taqqu, M.S., Springer pp. 55–77.

de Haan, L., 1994, Extreme value statistics, *In: Extreme Value Theory and Applications*, eds.: Galambos et al. pp. 93–122.

de Haan, L., and S.I. Resnick, 1993, Estimating the Limit Distribution of Multivariate Extremes, *Communications in Statistics - Stochastic Models* 9, 275–309.

de Haan, L., and U. Stadtmüller, 1996, Generalized regular variation of second order, *Journal of the Australian Mathematical Society* 61, 381–395.

Deheuvels, P., 1979, La fonction de dépendance empirique et ses propriétés, *Academia Royal of Belgium, Bull. C1 Sci. 5ième sér* 65, 274–292.

Deheuvels, P., 1981, A nonparametric test for independence, *Publications of the Institute of Statistics at the University of Paris* 26, 29–50.

Dowd, K., 1998, *Beyond Value at Risk: The New Science of Risk Management.* (John Wiley & Sons, Chichester).

Draisma, G., H. Drees, A. Ferreira, and L. de Haan, 2001, Tail dependence in independence, *Eurandom preprint.*

Drees, H., and E. Kaufmann, 1998, Selecting the optimal sample fraction in univariate extreme value estimation, *Stochastic Processes and their Applications* 75, 149–172.

Dudley, R. M., 1966, Weak convergence of measures on nonseparable metric spaces and empirical measures on Euclidean spaces, *Illinois Journal of Mathematics* 10, 109–126.

Dudley, R. M., 1967, Measures on nonseparable metric spaces, *Illinois Journal of Mathematics* 11.

Durrett, R., 1996, *Probability Theory and Examples.* (Duxbury Press) 2. edn.

Durrleman, V., A. Nikeghbali, and T. Roncalli, 2000, Which copula is the right one?, *Groupe de Recherche Opérationnelle, Crédit Lyonnais, preprint.*

Eberlein, E., 2001, Application of generalized hyperbolic Lévy motions to finance, *In: Lévy Processes: Theory and Applications,* eds.: Barndorff-Nielsen, T. Mikosch, and S. Resnick.

Eberlein, E., and U. Keller, 1995, Hyperbolic distributions in finance, *Bernoulli* 1, 281–299.

Eberlein, E., and K. Prause, 1999, The generalized hyperbolic model: financial derivatives and risk measures, Freiburg Center for Data Analysis and Modelling, University of Freiburg, preprint no. 56,.

Einmahl, J.H.J., L. de Haan, and V.I. Piterbarg, 2001, Multivariate Extremes Estimation, *Annals of Statistics* 19, 1401–1423.

Einmahl, J.H.J., L. de Haan, and A. K. Sinha, 1997, Estimating the spectral measure of an extreme value distribution, *Stochastic Processes and their Applications* 70, 143–171.

Embrechts, P., C. Klüppelberg, and T. Mikosch, 1997, *Modelling Extremal Events for Insurance and Finance.* (Springer Verlag, Berlin).

Embrechts, P., F. Lindskog, and A. McNeil, 2001, Modelling Dependence with Copulas and Applications to Risk Management, *ETH Zürich, preprint.*

Embrechts, P., A. McNeil, and D. Straumann, 1999, Correlation and Dependency in Risk Management: Properties and Pitfalls, *In: Risk Management: Value at Risk and Beyond,* ed.: M.A.H. Dempster, Cambridge University Press, Cambridge, pp. 176–223.

Embrechts, P., A. McNeil, and D. Straumann, 2001, Correlation and dependency in risk management: Properties and pitfalls, *In: Risk Management: Value at Risk Beyond* eds.: M. Dempster, and H.K. Moffat, Cambridge University Press.

Falk, M., J. Hüsler, and R.D. Reiss, 1994, *Laws of Small Numbers: Extremes and Rare Events.* (Birkhäuser, Basel).

Fang, K.T., S. Kotz, and K.W. Ng, 1990a, *Symmetric Multivariate and Related Distributions.* (Chapman and Hall London).

Fang, K.-T., S. Kotz, and K.-W. Ng, 1990b, *Symmetric multivariate and related distributions.* (Chapman & Hall, London).

Fang, K.-T., and Y.-T. Zhang, 1990, *Generalized multivariate analysis.* (Science Press, Beijing; Springer Verlag, Heidelberg).

Fermanian, J-D., D. Radulović, and M. Wegkamp, 2002, Weak convergence of empirical copula processes, *preprint.*

Fisher, R.A., and L.H.C.Tippett, 1928, Limiting forms of the frequency distributions of the largest or smallest member of a sample, *Proceedings of the Cambridge Philosophical Society* 24, 180–190.

Fletcher, Roger, 1987, *Practical Methods of Optimization.* (Wiley, New York).

Frahm, G., and M. Junker, 2003, Generalized elliptical distributions: Models and estimation, *Caesar Institute Bonn, preprint.*

Frahm, G., M. Junker, and A. Szimayer, 2003, Elliptical Copulas: Applicability and Limitations, *forthcoming in Statistics and Probability Letters.*

Freitag, E., and R. Busam, 1993, *Funktionentheorie 1.* (Springer Verlag New York).

Genest, C., K. Ghoudi, and L-P. Rivest, 1995, A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions, *Biometrika* 82, 543–552.

Genest, C., and L-P. Rivest, 1993, Statistical inference procedures for bivariate Archimedean copulas, *Journal of the American Statistical Association* 88, 1034–1043.

Gnedenko, B.V., 1943, Sur la distribution limité du terme d'une séries aléatoire, *Annals of Mathematics* 44, 423–453.

Gordy, M., and E. Heitfield, 2000, Estimating factor loadings when rating performance data are scarce, *Technical Report, Board of Governors of the Federal Reserve System.*

Gordy, M. B., 2001, A Risk-Factor Model Foundation for Ratings-Based Bank Capital Rules, *Board of Governors of the Federal Reserve System, preprint.*

Hahn, M.G., D.M. Mason, and D.C. Weiner , 1991, *Sums, trimmed sums and extremes.* (Birkhäuser, Boston).

Halgreen, C., 1979, Self-decomposatibility of the generalized inverse Gaussian and hyperbolic distribution functions, *Zeitung für Wahrscheinlichkeitstheorie* 47, 13–18.

Härdle, W., 1990a, *Applied nonparametric regression* vol. 19 of *Econometric Society Monographs.* (Cambridge University Press).

Härdle, W., 1990b, *Smoothing techniques: With implementation in S.* (Springer Verlag, Berlin).

Härdle, W., T. Kleinow, and G. Stahl, 2002, *Applied Quantitative Finance - Theory and Computational Tools.* (Springer Verlag, Berlin).

Hartmann, P., S. Straetmans, and C.G. de Vries, 2001, Asset market linkages in crisis periods, *Tinbergen Institute, preprint No 01-071/2.*

Hauksson, H.A., M. Dacorogna, T. Domenig, U. Mueller, and G. Samorodnitsky, 2001, Multivariate extremes, aggregation and risk estimation, *Quantitative Finance* 1, 79–95.

Heuser, H., 2000, *Lehrbuch der Analysis* vol. 2. (B.G. Teubner, Stuttgart) 11 edn.

Hodgson, D.J., O. Linton, and K. Vorkink, 2000, Testing the capital asset pricing model efficiently under elliptical symmetry: a semi-parametric approach, *London School of Economics, preprint.*

Hrycej, T., R. Schmidt, and E. Stützle, 2002, Multidimensional data modelling with generalized hyperbolic distributions, *preprint, www.mathematik.uni-ulm.de/finmath/.*

Huang, X., 1992, *Statistics of Bivariate Extreme Values.* (Thesis Publishers and Tinbergen Institute).

Hult, H., and F. Lindskog, 2002, Multivariate Extremes, Aggregation and Dependence in Elliptical Distributions, *Advances in Applied Probability 34(3)* pp. 587–609.

Joe, H., 1997, *Multivariate Models and Dependence Concepts.* (Chapman and Hall, London).

Jones, M.C., 1993, Simple boundary correction for kernel density estimation, *Statistics and Computing* 3, 135–146.

Jorion, P., 2000, *Value at Risk.* (McGraw-Hill, New York) 2nd. edn.

Junker, M., and A. May, 2002, Measurement of aggregate risk with copulas, *Research Center caesar, Germany, preprint.*

Karolyi, G.A., and R.M. Stulz, 1996, Why Do Markets Move Together? An Investigation of U.S.-Japan Stock Return Comovements, *Journal of Finance* 51, 951–989.

King, M., and S. Wadhwani, 1990, Transmission of volatility between stock markets, *The Review of Financial Studies* 3, 5–330.

Korsholm, L., 2000, The semi-parametric normal mean-variance mixture model, *Scandinavian Journal of Statistics* 27, 227–261.

Kullback, S, 1959, *Information theory and statistics.* (Wiley, New York).

Ledford, A.W., and J.A. Tawn, 1996, Statistics for near independence in multivariate extreme values, *Biometrika* 83, 169–187.

Ledford, A.W., and J.A. Tawn, 1997a, Concomitant tail behavior for extremes, *Advances in Applied Probability* 30, 197–215.

Ledford, A.W., and J.A. Tawn, 1997b, Modelling dependence within joint tail regions, *Journal of the Royal Statistical Society* 59, 475–499.

Li, R., K.-T. Fang, and L. Zhu, 1997, Some Q-Q probability plots to test spherical and elliptical symmetry, *Journal of Computational and Graphical Statistics* 6, 435–450.

Lindskog, F., 1999, Modelling Dependence with Copulas, *Risklab, Switzerland, Retrieved 2001-08-12 from http://www.risklab.ch/Papers.html#LCELindskog, preprint.*

Lindskog, F., 2000, Linear correlation estimation, *Risklab, Switzerland, Retrieved 2002-09-06 from http://www.risklab.ch/Papers.html#LCELindskog, preprint.*

Lindskog, F., A. McNeil, and U. Schmock, 2001, Kendall's tau for elliptical distributions, *ETH Zürich, preprint.*

Longin, F., and B. Solnik, 2001, Extreme Correlation of International Equity Markets, *The Journal of Finance* LVI, 649–676.

Loretan, M., 2000, Evaluating changes in correlations during periods of high market volatility, *Global Investor* 135, 65–68.

Madan, D.B., and E. Seneta, 1990, The Variance-Gamma (VG) model for share market returns, *Journal of Business* pp. 511–524.

Magnus, W., F. Oberhettinger, and R.P. Soni, 1966, *Formulas and Theorems for the Special Functions of Mathematical Physics* vol. 52 of *Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen.* (Springer Verlag, Berlin).

Malevergne, Y., and D. Sornette, 2002, Investigating extreme dependencies: concepts and tools, *University of Nice, Retrieved 2003-04-08 from http://papers.ssrn.com/sol3/papers.cfm?abstract id=303465, preprint.*

Manzotti, A., F.J. Pérez, and A.J. Quiroz, 2002, A statistic for testing the null hypothesis of elliptical symmetry, *Journal of Multivariate Analysis* 81, pp. 274–285.

Markowitz, H.M., 1987, *Mean-Variance Analysis in Portfolio Choice and Capital Markets.* (Basil Blackwell, Oxford).

Morgenstern, O., 1959, *International Financial Transactions and the Business Cycle.* (Princeton University Press, Princeton (NJ)).

Müller, H.G., 1991, Smooth optimum kernel estimators near endpoints, *Biometrika* 78, 521–530.

Nelsen, R.B., 1999, *An Introduction to Copulas.* (Springer Verlag, New York).

Neuhaus, G., 1971, On weak convergence of stochastic processes with multidimensional time paramter, *Annals of Mathematical Statistics* 42, 1285–1295.

Niederreiter, H., 1993, *Random number generation and quasi-Monte Carlo methods.* (SIAM Philadelphia).

Ong, M.K., 1999, *Internal Credit Risk Models.* (Risk Books, Haymarket).

Pagan, A.R., 1996, The econometrics of financial markets, *Journal of Empirical Finance* 3, 15–102.

Park, B.U., and J.S. Marron, 1990, Comparison of data-driven bandwidth selectors, *Journal of the American Statistical Association* 85, 66–73.

Parzen, E., 1962, On estimation of probability density function and mode, *Annals of Mathematical Statistics* 35, 1065–1076.

Peng, L., 1998, *Second Order Condition and Extreme Value Theory* vol. 178 of *PhD thesis, Tinbergen Institute Research Series.* (Thesis Publishers and Tinbergen Institute).

Peng, L., 1999, Estimation of the coefficient of tail dependence in bivariate extremes, *Statistics and Probability Letters* 43, 399–409.

Pickands, J., 1975, Statistical inference using extreme order statistics, *Annals of Statistics* 3, 119–131.

Pickands, J., 1981, Multivariate extreme value distributions, *In Proc. 43rd Session International Statistical Institute, Buenos Aires.*

Pollard, D., 1984, *Convergence of Stochastic Processes.* (Springer Verlag, New-York).

Prause, K., 1999, The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures, PhD Thesis, Albert-Ludwigs-Universität Freiburg i. Br.

Qi, Y., 1997, Almost sure convergence of the stable tail empirical dependence function in multivariate extreme statistics, *Acta Mathematicae Applicatae Sinica (English series)* 13, 167–175.

Rachev, S., and S. Mittnik, 2000, *Stable Paretian Models in Finance.* (Wiley, New York).

Ramchand, L., and R. Susmel, 1998, Volatility and cross correlation across major stock markets, *Journal of Empirical Finance* 5, 397–416.

Resnick, S., 1987, *Extreme Values, Regular Variation, and Point Processes.* (Springer, New York).

Resnick, S., 2002, Hidden Regular Variation, Second Order Regular Variation and Asymptotic Dependence, *preprint.*

Rinnooy Kan, A.H.G., and G.T. Timmer, 1987, Stochastic global optimization methods, Part I: Clustering methods, Part II: Multi-level methods, *Mathematical Programming* 39, 26–78.

RiskMetrics, 1995, Technical document, *JP. Morgan, New York.*

Ross, S.M., 1997, *Introduction to Probability models.* (Academic Press, San Diego) 6 edn.

Rousseeuw, P. J., and B.C. van Zomeren, 1990, Unmasking multivariate outliers and leverage points, *Journal of American Statistical Association* 85, 633–639.

Rüschendorf, 1976, Asymptotic normality of multivariate rank order statistics, *Annals of Statistics* 4, 912–923.

Ruymgaart, F.H., and J.A. Wellner, 1982, *Growth properties of multivariate empirical processes.* (Mathematisch Instituut Katholieke Universiteit, Nijmegen).

Ruymgaart, F. H., 1974, Asymptotic normality of nonparametric tests for independence, *Annals of Statistics* 2, 892–910.

Ruymgaart, F. H., G. R. Shorack, and W. R. van Zwet, 1972, Asymptotic normality of nonparametric test for independence, *Annals of Mathematical Statistics* 43, 1122–1135.

Sato, K.-I., 1999, *Lévy processes and infinite divisibility* vol. 68 of *Cambridge studies in advanced mathematics.* (Cambridge University Press).

Schmidt, R., 2002, Tail dependence for elliptically contoured distributions, *Mathematical Methods of Operations Research* 55, 301–327.

Schmidt, R., 2003a, Credit risk modelling and estimation via elliptical copulae, *In: Credit Risk: Measurement, Evaluation and Management,* eds.: G. Bohl, S. Rachev, and K.-H. Vollmer, Physica-Verlag Heidelberg.

Schmidt, R., 2003b, Tail dependence, *In: Statistical case studies,* eds.: P. Cizek and W. Härdle, Xplore, e-books, Springer Verlag.

Schmidt, R., and U. Stadtmüller, 2002, Nonparametric estimation of tail dependence, *submitted, www.mathematik.uni-ulm.de/finmath/.*

Schmitt, F., D. Schertzer, and S. Lovejoy, 1999, Multifractal analysis of foreign exchange data, *Application of Stochastic Models and Data Analysis* 15, 29–53.

Schoenberg, I.J., 1938, Metric spaces and completely monotone functions, *Annals of Mathematics* 41, 811–41.

Schuster, E.F., 1985, Incorporating support constraints into nonparametric estimators of densities, *Communications in Statistics and Theoretical Mathematics* 14, 1123–1136.

Schweizer, B., 1991, *Thirty years of copulas*, eds: G. Dall'Aglio, S. Kotz and G. Salinetti, *In: Advances in Probability Distributions with given Marginals,* Kluwer, Dordrecht.

Schweizer, B., and E.F. Wolff, 1981, On nonparametric measures of dependence for random variables, *Annals of Statistics* 9, 870–885.

Scott, D.W., 1992, *Multivariate density estimation: Theory, practice, and visualization.* (Wiley, New York).

Sheather, S.J., and M.C. Jones, 1991, A reliable data-based bandwidth selection method for kernel density estimation, *Journal of the Royal Statistical Society B* 53, 683–690.

Silvapulle, P., and C.W.J. Granger, 2001, Large returns, conditional correlation and portfolio diversification: a value-at-risk approach, *Quantitative Finance* 1, 542–551.

Sklar, A., 1996, Random variables, distribution functions, and copulas - a personal look backward and forward, *In: Distributions with Fixed Marginals and Related Topics,* eds.: L. Rueschendorff, B. Schweizer, and M.D. Taylor, pp. 1–14.

Skorokhod, A. V., 1956, Limit theorems for stochastic processes, *Theory of Probability and Its Applications* 1, 261–290.

Stărică, C., 1999, Multivariate extremes for models with constant conditional correlation, *Journal of Empirical Finance* 6, 515–553.

Stuart, Alan, and Keith Ord, 1994, *Kendall's Advanced Theory of Statistics, Volume I: Distribution Theory.* (Arnold, London).

Stute, W., 1984, The oscillation behavior of empirical processes: The multivariate case, *The Annals of Probability* 12, 361–379.

Stute, W., and U. Werner, 1991, Nonparametric estimation of elliptically contoured densities, *In: Nonparametric functional estimation and related topics,* eds.: G. Roussas, Kluwer Academic Publishers pp. 173–190.

Stützle, Eric A., and Tomas Hrycej, 2001, Forecasting of Conditional Distributions - An Application to the Spare Parts Demand Forecast, in *Proc. 2001 IASTED Int. Conf. on Artificial Intelligence and Soft Computing* Cancun, Mexico.

Stützle, Eric A., and Tomas Hrycej, 2002a, Estimating Multivariate Conditional Distributions via Neural Networks and Global Otimization, in *Proc. 2001 IEEE Int. Joint Conference on Neural Networks* Honolulu, Hawaii.

Stützle, Eric A., and Tomas Hrycej, 2002b, Modelling Future Demand by Estimating the Multivariate Conditional Distribution via the Maximum Likelihood Pronciple and Neural Networks, in *Proc. 2002 IASTED Int. Conf. on Modelling, Identification and Control* Innsbruck, Austria.

Susmel, R., and R.F. Engle, 1994, Hourly volatility spillovers between international equity markets, *Journal of International Money and Finance* 13, 3–25.

Tawn, J., 1988, Bivariate extreme value theory, *Biometrika* 75, 397–415.

Van der Vaart, A. W., and J. A. Wellner, 1996, *Weak Convergence and Empirical Processes.* (Springer Verlag, New York).

Venables, W.N., and B.D. Ripley, 1999, *Modern Applied Statistics with S-PLUS.* (Springer Verlag, New York) 3rd edn.

Wand, M.P., and M.C. Jones, 1995, *Kernel Smoothing.* (Chapman & Hall, London).

Wheeden, R.L., and A. Zygmund, 1977, *Measure and Integral - An Introduction to real analysis.* (Marcel Dekker Inc.).

Whittaker, E. T., and G. N. Watson, 1996, *A course of modern analysis.* (Cambridge University Press Cambridge).

List of Tables

List of Figures

Index

List of Symbols and Abbreviations

$\mathbb{R}^n$	n-dimensional real space $(-\infty, \infty)^n$	15
$\mathbb{R}^n_+$	n-dimensional positive real space $[0, \infty)^n$	15
$\bar{\mathbb{R}}^n_+$	n-dimensional extended real space $[0, \infty]^n \backslash \{(\infty, \ldots, \infty)\}$	25, 70
$\mathbb{R}^{n \times n}$	space of real $n \times n$-matrices	32
$l^\infty(T)$	space of uniformly bounded, real functions on some set T	77
$C(T)$	space of continuous functions on some set T	83
$\mathcal{B}_\infty(\bar{\mathbb{R}}^n_+)$	space of functions $f : \bar{\mathbb{R}}^n_+ \to \mathbb{R}$ which are locally uniformly-bounded on every compact subset of $\bar{\mathbb{R}}^n_+$	77
$\mathbb{S}^{n-1}$	$(n-1)$-dimensional unit-sphere $\{x \in \mathbb{R}^n : \lVert x \rVert = 1\}$	29
$\overline{\mathbb{S}}^{n-1}$	space $\mathbb{S}^{n-1} \backslash (-\infty, 0]^n$	15
RV_α	class of regularly varying functions with index α	27
OR	class of O-regularly varying functions	27
Φ_n	characteristic generator	32
$S_n(\Phi)$	family of n-dimensional spherically contoured distributions with characteristic generator Φ_n	32
$E_n(\mu, \Sigma, \Phi)$	family of n-dimensional elliptically contoured distributions with parameters μ, Σ, and characteristic generator Φ_n	47
MGH_n	family of n-dimensional generalized hyperbolic distributions	123
$MAGH_n$	family of n-dimensional affine generalized hyperbolic distributions	124
$C(u)$	copula function $C : [0, 1]^n \to [0, 1]$	21
λ_U	upper tail-dependence coefficient (TDC)	23
λ_L	lower tail-dependence coefficient (TDC)	23
$\Lambda_U^{I,J}$	upper tail-copula	25
$\Lambda_L^{I,J}$	lower tail-copula	25

$\hat{\lambda}_{U,m}$	nonparametric estimator for the upper TDC	73
$\hat{\lambda}_{L,m}$	nonparametric estimator for the lower TDC	73
$\hat{\lambda}_{U,m}^{EVT}$	modified nonparametric estimator for the upper TDC	73
$\hat{\lambda}_{L,m}^{EVT}$	modified nonparametric estimator for the lower TDC	73
$\hat{\lambda}_{m}^{E}$	parametric estimator for the TDC of elliptical distributions	92
$\hat{\Lambda}_{U,m}$	upper empirical tail-copula	72
$\hat{\Lambda}_{L,m}$	lower empirical tail-copula	72
$\hat{\Lambda}_{U,m}^{EVT}$	modified upper empirical tail-copula	73
$\hat{\Lambda}_{U,m}^{*}$	upper empirical tail-copula under known margins	78
$\hat{\Lambda}_{L,m}^{*}$	lower empirical tail-copula under known margins	78
$\mathbb{G}$	Gaussian random field	79
$\check{k}$	Mellin transform	28
$g \overset{M}{*} f$	Mellin convolution between the functions g and f	28
ϕ_n	characteristic function or generator of an Archimedean copula	32,65
K_ν	modified Bessel function of the third kind or MacDonald function	54
$\|\| \cdot \|\|$	norm on some space	29
M'	transpose of a matrix M	32
$\overset{v}{\rightarrow}$	vague convergence of measures	28
$\overset{w}{\rightarrow}$	weak convergence of measures	76
TDC	tail-dependence coefficient	22
EVT	extreme-value-theory	12
VaR	Value-at-Risk	13
MGH	multivariate generalized hyperbolic (distribution)	123
MAGH	multivariate affine generalized hyperbolic (distribution)	124
RMSE	root mean - squared error	114
MESE	mean error - standard error	114
RBP	risk-based pricing	177

Zusammenfassung

Im Mittelpunkt der vorliegenden Dissertation steht die Modellierung und die statistische Untersuchung von Abhängigkeiten extremer Ereignisse. Diverse Simulationsstudien und Datenanalysen unterstreichen die Anwendbarkeit der entwickelten Konzepte und Resultate im Bereich der Finanzwirtschaft. Darüber hinaus werden mehrere integrative Verteilungsmodelle zur realistischeren Modellierung von Finanzdaten vorgeschlagen.

Hinsichtlich der Modellierung von multidimensionalen extremen Ereignissen erweist sich die multidimensionale Extremwerttheorie bzw. die multivariaten Extremwertverteilungen als besonders geeignet. Jedoch sind die Abhängigkeitsstrukturen multivariater Extremwertverteilungen und deren Zusammenhang zur Wahrscheinlichkeitsverteilung der Gesamtstichprobe den meisten Theoretikern und Praktikern unbekannt. Eines der Ziele dieser Dissertation ist deshalb die Beschreibung und Modellierung von sogenannten *Tail-Copula*-Funktionen, welche die Abhängigkeitsstruktur von multivariaten Extremwertverteilungen in geeigneter Weise beschreiben und sich gut in die Theorie der gewöhnlichen Copulae (oder Abhängigkeitsfunktionen) eingliedern. Als allgemeines Abhängigkeitsmaß für extreme Ereignisse scheint die Klasse der Tail-Copula-Funktionen jedoch aufgrund ihrer Komplexität in der Praxis ungeeignet zu sein. Der durchschlagende Erfolg des Korrelationskoeffizienten in finanzwirtschaftlichen Anwendungen, als Abhängigkeitsmaß zwischen bivariaten Verteilungen, geht nicht zuletzt darauf zurück, dass die Abhängigkeit lediglich durch einen einzigen Parameter quantifiziert wird. Aus diesem Grund konzentriert sich die vorliegende Arbeit besonders auf die sogenannten oberen und unteren *Tail-Abhängigkeitskoeffizienten*, welche die Abhängigkeit von extremen Ereignissen in geeigneter Weise messen. In diesem Zusammenhang entwickeln wir grundlegende Charakterisierungen und Schätzmethoden für eine große Anzahl von Verteilungen und Copulae. Die Klasse der multivariaten elliptisch-konturierten Verteilungen bzw. Copulae spielt dabei in dieser Dissertation eine herausragende Rolle. Diese Klasse beinhaltet die Familie der Normalverteilungen bzw. deren Copulae als Spezialfall, und viele Eigenschaften der Normalverteilung lassen sich auf allgemeine elliptisch-konturierte Verteilungen übertragen. Zum Beispiel existieren gerade in höheren Dimensionen schnelle Generatoren für entsprechende Pseudo-Zufallsvektoren und robuste Schätzalgorithmen für die Verteilungsparameter. Für finanzwirtschaftliche Anwendungen stellt die Familie der elliptisch-konturierten Verteilungen somit eine interessante Alternative zur Klasse der Normalverteilungen dar.

Ergänzend zur Untersuchung von Abhängigkeitsstrukturen extremer Ereignisse be-

trachten wir verschiedene integrative Verteilungsmodelle, deren ganzheitliche Abhängigkeitsstruktur auf die Modellierung von Finanzmarkdaten zugeschnitten ist. Hierbei konzentrieren wir uns insbesondere auf die Modellierung von Returns diverser Finanztitel. Moderne Markt- und Kreditrisikomodelle verwenden integrative multivariate Verteilungsfunktionen, wie zum Beispiel die Normalverteilungen (vgl. RiskMetrics (1995), CreditMetrics (1998)), verallgemeinerte hyperbolische Verteilungen (vgl. Bingham and Kiesel (2001a), Eberlein (2001)) oder alpha-stabile Verteilungen (vgl. Rachev and Mittnik (2000)) zur Modellierung der Returns von Finanzpapieren. Eine große Anzahl empirischer Studien zeigt jedoch (vgl. Abschnitt 2.1.3 und Kapitel 5), dass die Normalverteilungen gerade im Risikomanagement oft unerwünschte Eigenschaften besitzen. Diese sind hauptsächlich auf die dünnen Tails und die starke Symmetrie der eindimensionalen Randverteilungen zurückzuführen. Andererseits führen flexiblere Verteilungsmodelle wie die verallgemeinerten hyperbolischen Verteilungen oder die alpha-stabilen Verteilungen häufig zu Schwierigkeiten hinsichtlich der Parameterschätzung, der Generierung entsprechender Pseudo-Zufallsvektoren oder der Modellierung geeigneter Abhängigkeitsstrukturen. Zur besseren Handhabung dieser Probleme stellen wir mehrere alternative Verteilungsmodelle vor, die einerseits auf multivariaten elliptisch-konturierten Verteilungen (eines speziellen semi-parametrischen Typs) und andererseits auf multivariaten affinen generalisiert-hyperbolischen Verteilungen basieren. Aus praktischer Perspektive besitzen beide Verteilungsklassen geeignete Abhängigkeitsstrukturen und gute statistische Eigenschaften, insbesondere für höherdimensionale Modellierungen.

Kapitel 2 stellt dem Leser die notwendigen Grundlagen aus der Extremwerttheorie bereit und führt die folgenden theoretischen Konzepte ein: Copulae, Tail-Copulae, Tail-Abhängigkeit, reguläre Variation, O-reguläre Variation und multidimensionale reguläre Variation. Zusätzlich stellen wir verschiedene visuelle statistische Hilfsmittel zur Erkennung von Abhängigkeiten extremer Ereignisse in Finanzdaten vor.

In Kapitel 3 untersuchen wir das Konzept der Tail-Abhängigkeit für verschiedene Verteilungen und Copulae. Die Abschnitte 3.1 and 3.2 enthalten eine allgemeine Charakterisierung für tail-abhängige sphärisch- und elliptisch-konturierte Verteilungen. Die Hauptresultate sind in den Theoremen 3.1.4 and 3.2.2 dargestellt. Insbesondere zeigen wir, dass sphärisch- und elliptisch-konturierte Verteilungen tail-abhängig sind, falls der Tail der zugehörigen generierenden Zufallsvariablen oder des entsprechenden Dichte-Generators regulär variiert. Letzteres heißt, dass der Tail essentiell durch eine Potenzfunktion approximiert werden kann. Weiterhin zeigen wir eine notwendige Bedingung für das Vorliegen von Tail-Abhängigkeit, welche etwas schwächer als die reguläre Variation des obigen Tails ist. Bei letzterem handelt es sich um O-regulär variierende Tails. Das Wachstum bzw. der Abfall von O-regulär variierenden Funktionen liegt dabei zwischen zwei Potenzfunktionen. Zusätzlich entwickeln wir eine explizite Formel (siehe Formel (3.18)) für den unteren bzw. oberen Tail-Abhängigkeitskoeffizienten von elliptisch-konturierten Verteilungen. Die genannten Charakterisierungen wurden in Schmidt (2002) veröffentlicht und bilden nach unserem Wissensstand neue Erkenntnisse. Die hinreichende Bedingung für die Tail-Abhängigkeit von elliptisch-konturierten Verteilungen wurde kürzlich ebenfalls in Hult and Lindskog (2002) unter einer anderen Beweisführung publiziert. Neben den bisher genannten Ergebnissen beweisen wir weitere interessante Resultate hinsichtlich der Beziehung zwischen regulär (O-regulär) variierenden generierenden Zufallsvariablen, Dichte-Generatoren und Randverteilungsfunktionen von sphärisch- und elliptisch-konturierten Verteilungen. Weiterhin überprüfen wir, ob bekannte elliptisch-konturierte Verteilungen wie die multivariaten Normal-, t-, logistische und symmetrischen generalisiert-hyperbolischen Verteilungen tail-abhängig oder tail-unabhängig sind.

In Abschnitt 3.3 wird das Konzept der Tail-Abhängigkeit für elliptisch-konturierte Verteilungen und Copulae in die multidimensionale Extremwerttheorie eingebettet. Insbesondere leiten wir in den Theoremen 3.3.2 und 3.3.5 hinreichende Bedingungen für elliptisch-konturierte Verteilungen und Copulae her, die im Anziehungsbereich von Extremwertverteilungen liegen. Letztendlich werden in Abschnitt 3.4 verschiedene Charakterisierungen von Tail-Abhängigkeit für Archimedische Copulae diskutiert und in Abschnitt 3.5 weitere Copulae in Bezug auf Tail-Abhängigkeit untersucht.

Kapitel 4 ist der Schätzung von Abhängigkeiten extremer Ereignisse gewidmet. In Abschnitt 4.1 wird das Konzept der Tail-Copulae eingeführt, welches die Abhängigkeitsstruktur im Tail von multivariaten Verteilungen beschreibt und welches eine Verallgemeinerung des Konzepts der Tail-Abhängigkeit darstellt. Mehrere nichtparametrische Schätzer für Tail-Copulae werden in Abschnitt 4.2 vorgestellt. Weiterhin schlagen wir einen nichtparametrischen Schätzer für den unteren bzw. oberen Tail-Abhängigkeitskoeffizienten vor. In Abschnitt 4.3 zeigen wir, dass der Begriff *Tail-Copula* seine Berechtigung hat, da sich viele Eigenschaften der gewöhnlichen Copula-Funktionen auf Tail-Copulae übertragen lassen. Um die asymptotische Normalität und starke Konsistenz der vorgestellten Schätzer zu diskutieren, führen wir in Abschnitt 4.4 den Funktionenraum der lokal gleichmäßig beschränkten Funktionen auf kompakten Mengen des $\bar{I\!R}_+^n$ ein. Unter Verwendung einer allgemeinen funktionalen Delta-Methode können wir in den Abschnitten 4.5 und 4.6 verschiedene Resultate über die asymptotische Normalität und die starke Konsistenz der obigen Schätzer beweisen. In diesem Zusammenhang erweist sich die Definition der schwachen Konvergenz bezüglich sogenannter äußerer Erwartungen als geeignet (siehe Van der Vaart and Wellner (1996)). Der Beweis des Hauptsatzes (vgl. Theorem 4.5.3) erfolgt in zwei Schritten. Zunächst beweisen wir in Theorem 4.5.2 die asymptotische Normalität für Verteilungen mit bekannten Randverteilungen. Im zweiten Schritt wird die Annahme der bekannten Randverteilungen fallen gelassen. Nach unserem Wissensstand stellen diese Resultate neue Erkenntnisse dar und bilden Erweiterungen der Resultate in Huang (1992); die Beweistechniken sind allerdings voneinander verschieden. In Abschnitt 4.7 untersuchen wir allgemeine Rangordnungsstatistiken von extremen Ereignissen und erarbeiten ein naheliegendes Resultat über die schwache Konvergenz dieser Statistiken. Ähnliche Betrachtungen für gewöhnliche Rangordnungsstatistiken wurden bereits von Ruymgaard, Shorack und van Zwet (1972), Ruymgaart (1974), Rüschendorf (1976) und Fermanian, Radulović und Wegkamp (2002) durchgeführt.

In Abschnitt 4.9 konzentrieren wir uns auf diverse Schätzproblematiken für elliptisch-konturierte Verteilungen. Insbesondere schlagen wir einen attraktiven parametrischen

Schätzer für den Tail-Abhängigkeitskoeffizienten vor, welcher auf der Formel (3.18) basiert. Weiterhin stellen wir einen robusten Schätzer für die Korrelationsmatrix vor, der auf die Familie der elliptisch-konturierten Verteilungen zugeschnitten ist. In Abschnitt 4.10 erarbeiten wir einen Leitfaden mit verschiedenen Schätzern des unteren bzw. oberen Tail-Abhängigkeitskoeffizienten, welche für beliebige Wahrscheinlichkeitsverteilungen anwendbar sind. Hierbei unterscheiden wir zwischen parametrischen und nichtparametrischen Schätzern sowie zwischen Schätzmethoden, die auf der gesamten Stichprobe basieren oder die lediglich extreme Realisationen in die Schätzung einbeziehen. Außerdem werden verschiedene statistische und empirische Eigenschaften der Schätzer vorgestellt und mögliche Anwendungsgebiete diskutiert. Abschließend vergleicht eine ausführliche Simulationsstudie in Abschnitt 4.11 die obigen Schätzer hinsichtlich ihrer Schätzcharakteristika und veranschaulicht diverse statistische Eigenschaften, die in den vorangegangen Abschnitten erarbeitet wurden.

In Kapitel 5 diskutieren wir verschiedene integrative Verteilungsmodelle, indem wir deren Abhängigkeitsstruktur, Parameterschätzung, Generierung von Pseudo-Zufallsvektoren und Anwendbarkeit zur Anpassung von Returns mehrerer Finanztitel untersuchen. In Abschnitt 5.1 beschäftigen wir uns zunächst mit Verteilungen, die verallgemeinerte hyperbolische Randverteilungen besitzen. Insbesondere betrachten wir in Abschnitt 5.1.1 multivariate verallgemeinerte hyperbolische Verteilungen, denen es bei hochdimensionalen Modellierungen häufig an robusten und schnellen Schätzmethoden fehlt. Aus diesem und aus anderen Gründen führen wir in Abschnitt 5.1.2 eine neue Klasse von Verteilungen ein, welche aus affin-linearen Transformationen von Zufallsvektoren mit unabhängigen und verallgemeinerten hyperbolisch verteilten Rändern hervorgehen. Gemäß Abschnitt 5.1.3 besitzen diese Verteilungen gute Schätzeigenschaften und haben eine attraktive Abhängigkeitsstruktur. In diesem Zusammenhang entwickeln wir in den Abschnitten 5.1.5 und 5.1.6 die notwendigen Algorithmen zur Parameterschätzung und zur Generierung von Pseudo-Zufallsvektoren. Bezüglich letzterem stellen wir einen Pseudo-Softwarecode für zwei effiziente Generatoren von Zufallsvektoren vor, welche auf einem Ablehnungsalgorithmus mit einfacher Hüllkurve basieren. Die Vorteile und Nachteile der beiden oben genannten Verteilungsklassen werden in den Abschnitten 5.1.7 und 5.1.8 diskutiert und mittels einer Simulationsstudie veranschaulicht.

Abschnitt 5.2 bezieht sich auf eine Familie von semi-parametrischen Verteilungen, welche im Großen und Ganzen eine Unterklasse der elliptisch-konturierten Verteilungen bildet. Insbesondere modellieren wir einerseits den Dichte-Generator einer elliptisch-konturierten Verteilung mittels nichtparametrischer Methoden, andererseits die sogenannte Skalierungsmatrix sowie den Lokationsvektor mittels parametrischer Methoden. Hierbei stellt die Klasse der (symmetrischen) verallgemeinerten hyperbolischen Verteilungen ein Spezialfall dar. Das Hauptaugenmerk liegt auf Normal-Varianz-Mixturen mit einer self-decomposable Mixturverteilung. In Abschnitt 5.2.4 passen wir die semi-parametrischen Verteilungen an diverse Finanzzeitreihen an und diskutieren den Value-at-Risk von linearen Wertpapierportfolien, die auf den obigen Verteilungsannahmen basieren.

Das letzte Kapitel dieser Dissertation, Kapitel 6, beschäftigt sich mit speziellen praktischen Anwendungen der vorangegangenen Verteilungsmodelle und Abhängigkeitsmaße. Abschnitt 6.1 stellt eine Anwendung der obigen Modelle und Techniken im Kontext eines risiko-basierten Kreditbewertungssystems vor. Dieses System wurde im Rahmen einer Kooperation mit der DaimlerChrysler AG, Forschung und Entwicklung Ulm, erfolgreich implementiert. In Abschnitt 6.2 diskutieren wir kurz das Portfoliomodell des Internal Ratings based Approach der Neuen Basler Eigenkapitalvereinbarung, welches zur Berechnung von regulatorischem Eigenkapital dient. Abschnitt 6.3 stellt weitere Resultate der obigen Kooperation vor. Abschließend wird in Abschnitt 6.4 der untere Tail-Abhängigkeitskoeffizient diverser Finanzzeitreihen über die Zeit hinweg geschätzt und einige Schlussfolgerungen hinsichtlich seiner Aussagekraft gezogen.